VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra aplikované matematiky

Benfordův zákon
Benford's Law

2016 Ondřej Markovič

VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra aplikované matematiky

Zadání bakalářské práce

Student: **Ondřej Markovič**

Studijní program: B2647 Informační a komunikační technologie

Studijní obor: 1103R031 Výpočetní matematika

Téma: Benfordův zákon
Benford's Law

Jazyk vypracování: čeština

Zásady pro vypracování:

Při studiu celé řady jevů se vyskytují realizace tzv. Benfordova pravděpodobnostního rozdělení. Cílem práce je popis tohoto rozdělení, studium podmínek vedoucích k jeho vzniku a přehled situací, kdy dochází k jeho uplatnění v praxi. Součástí práce je popis a ilustrace statistického ověřování hypotézy o tom, že vyšetřovaná data pocházejí z Benfordova rozdělení.

Seznam doporučené odborné literatury:

Anděl, J.: Základy matematické statistiky, Matfyzpress, Praha, 2007
Hill, T. P.: A Statistical Derivation of the Significant-Digit Law, statistical Science 1995,Vol. 10, No. 4, 354-363

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **Mgr. Bohumil Krajc, Ph.D.**

Datum zadání: 01.09.2015
Datum odevzdání: 29.04.2016

Bouchala

doc. RNDr. Jiří Bouchala, Ph.D.
vedoucí katedry

VYSOKÁ ŠKOLA BÁŇSKÁ-TECHNICKÁ UNIVERZITA
Fakulta elektrotechniky a informatiky
1D16
OSTRAVA

prof. RNDr. Václav Snášel, CSc.
děkan fakulty

„Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.“

V Ostravě 29. dubna 2016

....

Děkuji panu Mgr. Bohumilovi Krajcovi, Ph.D. za pomoc při tvorbě práce.

Abstrakt

Cílem této práce je představení fenoménu zvaného Benfordův zákon. Benfordův zákon popisuje specifické, nerovnoměrné rozdělení signifikantních číslic v různých číselných matematických strukturách i v souborech přirozeně se vyskytujících čísel. Rozdělení signifikantních číslic umožňuje, porovnáváním číselných souborů s Benfordovým zákonem, odhalit například volební i účetní podvody a tímto způsobem být dokonce použito i jako důkaz.

V této práci popíšeme rozdělení signifikantních číslic dané Benfordovým zákonem a definujeme některé benfordovské matematické struktury. Dále rozebereme specifické vlastnosti, které nám umožní nalézt číselné soubory vyhovující Benfordovu zákonu. Krátce se zmíníme o diferenciálních rovnicích v souvislosti s Benfordovým zákonem a pokusíme se o vysvětlení výskytu Benfordova zákona v mnoha číselných souborech. V závěru prezentujeme přirozeně se vyskytující číselné soubory, které porovnáváme s Benfordovým zákonem.

Klíčová slova: Benfordův zákon, Benfordovo pravděpodobnostní rozdělení, benfordovské matematické struktury, signifikantní číslice, funkce signifikandu, rovnoměrné rozdělení modulo 1, invariance rozdělení signifikantních číslic vzhledem ke změně měřítka.

Abstract

The goal of this work is to present a phenomenon called Benford's Law. This law describes a unique non-uniform distribution of significant digits in various numerical mathematical objects and naturally occurring datasets. Distribution of significant digits in real life allows, by comparing collections of numbers with Benford's Law, to expose for example election and accounting frauds and in this way be even used as evidence.

In this thesis, we will describe this unique distribution of Benford's Law and define some benford mathematical objects. Next we will analyze many unique properties, that will allow us to find collections of numbers that are following Benford's Law. We will discuss differential equations in context with Benford's Law and we will try to explain the occurance of Benford's Law in many collections of numbers. In the end we will present naturally occurring datasets, which we compare to Benford's Law.

Keywords: Benford's Law, Benford's probability distribution, benford mathematical objects, significant digits, significand function, uniform distribution modulo 1, scale-invariance.

Obsah

Seznam symbolů

Symboly	**Popis symbolů**
$A \setminus B$	prvky A které nejsou v B
$\mathcal{B}$	borelovské sigma algebry
$\mathbb{B}$	Benfordovo pravděpodobnostní rozdělení
C^k	pro $k \in \{\mathbb{N}, 0\}$množina všech funkcí, které jsou k-krát spojitě diferencovatelné
D_m	pro $m \in \mathbb{N}$ označuje m-tou signifikantní číslici
EX	střední hodnota náhodné veličiny X
$\mathcal{F}(x)$	označuje zlomkovou část x
F_X	distribuční funkce náhodné veličiny X
f_X	pravděpodobnostní hustota náhodné veličiny X
f^{-1}	vzor funkce f
$f_*\mathbb{P}$	indukovaná pravděpodobnostní míra
$\mathbb{I}_{\langle a,b)}$	charakteristická funkce intervalu $\langle a, b)$
$i.r.s.m.$	invarience rozdělení signifikantních číslic vzhledem ke změně měřítka
$\mathbb{N}$	přirozená čísla
O	velké O
$\mathbb{P}$	označení pravděpodobnostní míry
$\mathbb{R}$	reálná čísla
$\mathscr{S}$	signifikand sigma algebra
S	funkce signifikandu
$s.j.$	skoro jistě, tedy s pravděpodobností 1
$X_n \stackrel{\mathcal{D}}{\to} X$	posloupnost náhodných veličin (X_n) konverguje v distribuci k X
$X_n \stackrel{s.j.}{\to} X$	posloupnost náhodných veličin (X_n) konverguje k X *skoro jistě*
$\mathbb{Z}$	celá čísla
δ_ω	Dirakova míra
λ	Lesbegueova míra
$\lambda_{a,b}$	symbol spojitého rovnoměrného rozdělení pravděpodobnosti na $\langle a, b)$
$\sigma(\mathcal{E})$	nejmenší sigma algebra generovaná $\mathcal{E}$
$(\Omega, \mathcal{A}, \mathbb{P})$	pravděpodobnostní prostor
$\#A$	mohutnost množiny A
$\lfloor x \rfloor$	dolní celá část x, čili největší celé číslo menší než x
$\lvert x \rvert$	absolutní hodnota z x
(x_n)	posloupnost čísel, $n \in \mathbb{N}$
(F_n)	posloupnost Fibonacciho čísel
(p_n)	posloupnost prvočísel
$\forall x$	oznamuje „pro všechna x“
$\exists x$	oznamuje „existuje x takové, že“
$\square$	konec Důkazu

Seznam obrázků

Seznam tabulek

Úvod

Benfordův zákon pojednává o specifickém logaritmickém rozdělení signifikantních číslic v mnoha číselných souborech. Vezmeme-li si například délky řek, počet obyvatel v obcích, seznamy fyzikálních a matematických konstant, pak číslice nebudou vyhovovat rovnoměrnému rozdělení, jak by se dalo očekávat. Místo toho se frekvence, s jakou se signifikantní číslice vyskytují v číselném souboru vyhovujícím Benfordovu zákonu, s „vyššími“ číslicemi snižuje. Tento zprvu značný odstup mezi jednotlivými četnostmi prvních signifikantních číslic s následujícími signifikantními číslicemi postupně mizí a místo toho se blíží k rovnoměrnému rozdělení.

Benfordův zákon je používán jako důkaz volebních i účetních podvodů. Užití Benfordova zákona tímto způsobem nejvíce popularizoval jihoafrický matematik Mark J. Nigrini. Snaha je také o užití tohoto zákona v numerických metodách, k analýze zaokrouhlovacích chyb.

V práci se z matematického hlediska budeme snažit o co nejlepší popsání Benfordova zákona. V první části popíšeme stručnou historii vzniku Benfordova zákona. Dále uvedeme signifikantní číslice a důvod jejich použití, a poté popíšeme logaritmické rozdělení signifikantních číslic, na kterém stojí samotný Benfordův zákon. Nakonec zkusíme porovnat dvě známé posloupnosti s Benfordovým zákonem.

Ve druhé části, užitím signifikandu, nadefinujeme funkci signifikandu, která nám pomůže hlavně při zápisu vícero signifikantních číslic.

V třetí části zkoumáme nejznámější matematické struktury v souvislosti s Benfordovým zákonem. To nám umožní porovnávat tyto matematické struktury s jinými, ale stejného typu, a tím nám poskytnout informaci, jestli jsou benfordovské.

Ve čtvrté části popíšeme specifické vlastnosti Benfordova pravděpodobnostního rozdělení, které nám usnadní hledání číselných souborů blízkých, či vyhovujících Benfordovu zákonu.

V následujících dvou částech stručně popíšeme diferenciální rovnice v souvislosti s Benfordovým zákonem a pokusíme se o vysvětlení výskytu Benfordova zákona v mnoha číselných souborech.

V poslední části aplikujeme informace získané z předchozích částí na dvě zmiňované posloupnosti a dále otestujeme Pearsonovým χ^2 testem dobré shody některé získané z praxe.

Teoretická část práce podstatným způsobem vychází z rozsáhlého článku [BH]. Pokud jsou v práci tvrzení převzatá z jiných zdrojů, je u každého z nich uvedena citace. Kromě několika málo vlastních jednoduchých příkladů jsem se snažil stručné informace z uvedené publikace co nejvíce rozepsat a komentovat. Originální část této práce je obsažena zejména v části věnované identifikaci benfordovských souborů a v připojeném software.

O míře a pravděpodobnosti

Následující část bude tvořena převážně definicemi nutnými k sjednocení terminologie z některých částí statistiky a pravděpodobnosti.

Definice 1. Nechť $\mathcal{A}$ je systémem podmnožin množiny Ω takovým, že platí:

$$\Omega \in \mathcal{A},$$

$$A \in \mathcal{A} \implies \Omega \setminus A \in \mathcal{A},$$

$$\forall n \in \mathbb{N}: \; A_n \in \mathcal{A} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.$$

Pak tento systém podmnožin nazveme *σ-algebrou na* Ω.

Definice 2. Buď $\mathcal{E}$ systémem podmnožin množiny Ω. Pak nejmenší σ-algebru na Ω, z hlediska uspořádání daného inkluzí, obsahující $\mathcal{E}$ nazveme *σ-algebrou generovanou* $\mathcal{E}$ a značíme $\sigma(\mathcal{E})$.

Poznámka. Nejmenší σ-algebra na Ω obsahující $\mathcal{E}$ je rovna průniku všech možných σ-algeber na Ω obsahujících $\mathcal{E}$.

Definice 3. *Borelovská σ-algebra* $\mathcal{B}(\mathbb{R})$ je σ-algebra na $\mathbb{R}$ generovaná všemi intervaly. Prvky borelovské σ-algebry nazveme borelovskými množinami.

Definice 4. Necht $C \subset \mathbb{R}$. Pak *Borelovská σ-algebra* $\mathcal{B}(C)$, je σ-algebra na C taková, že platí:

$$\mathcal{B}(C) = C \cap \mathcal{B}(\mathbb{R}) = \{C \cap B : \; B \in \mathcal{B}(\mathbb{R})\}.$$

Poznámka. V práci budou zvoleny kratší formy zápisu borelovské σ-algebry. Tedy $\mathcal{B}$ místo zápisu $\mathcal{B}(\mathbb{R})$, $\mathcal{B}\langle a, b)$ je borelovská σ-algebra na $\langle a, b)$ a $\mathcal{B}^+$ bude označovat $\mathcal{B}(\mathbb{R}^+)$.

Definice 5. Nechť je dána funkce $f : \Omega \to \mathbb{R}$. Pak množinu

$$f^{-1}(C) = \{\omega \in \Omega : \; f(\omega) \in C\},$$

nazveme *vzorem množiny* C při zobrazení f.

Definice 6. Nechť je dána funkce $f : \Omega \to \mathbb{R}$. Pak σ-algebru danou předpisem

$$\sigma(f) = \{f^{-1}(B) : \; B \in \mathcal{B}\},$$

nazveme σ-algebrou generovanou funkcí f.

Definice 7. Nechť je dán spočetný systém množin $A_1, A_2 \ldots A_n \ldots$ z $\mathcal{A}$. Jestliže platí

$$\forall i = \mathbb{N}, \forall j = \mathbb{N}, i \neq j : \; A_i \cap A_j = \emptyset,$$

potom řekneme, že prvky z tohoto spočetného systému množin jsou *po dvou disjunktní.*

Definice 8. *Mírou* na $(\Omega, \mathcal{A})$, kde Ω je neprázdná množina a $\mathcal{A}$ je σ-algebra na Ω, nazveme funkci $\mu : \mathcal{A} \to \langle 0, \infty)$, pokud $\mu(\emptyset) = 0$ a pro spočetný systém množin $A_1, A_2 \ldots A_n \ldots$ z $\mathcal{A}$, které jsou po dvou disjunktní, platí

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

Definice 9. *Lebesgueovou mírou* nazveme míru λ na $(\mathbb{R}, \mathcal{B})$, jestliže pro každý interval $\langle a, b) \subset \mathbb{R}$ platí:

$$\lambda(\langle a, b)) = b - a.$$

Definice 10. *Borelovsky měřitelnou funkcí* nazveme funkci $f : \mathbb{R} \to \mathbb{R}$, pokud pro libovolný interval $I \subset \mathbb{R}$ platí:

$$f^{-1}(I) \in \mathcal{B}.$$

Definice 11. *Pravděpodobnostní mírou* na $(\Omega, \mathcal{A})$, kde Ω je neprázdná množina a $\mathcal{A}$ je σ-algebra na Ω, nazveme funkci $\mathbb{P} : \mathcal{A} \to \langle 0, 1\rangle$, pro kterou platí:

$$\mathbb{P}(\Omega) = 1,$$

a pro každý spočetný systém po dvou disjunktních množin $A_1, A_2 \dots A_n \dots$ z $\mathcal{A}$ platí:

$$\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Poznámka. Jinými slovy, pravděpodobnostní míra je takovou mírou $\mathbb{P}$, pro kterou platí $\mathbb{P}(\Omega) = 1$. Pro pravděpodobnostní míru se může používat i označení *pravděpodobnost*, *rozdělení*, nebo *pravděpodobnostní rozdělení*. Borelovskou pravděpodobnostní mírou nazýváme pravděpodobnostní míru na $(\mathbb{R}, \mathcal{B})$.

Definice 12. *Pravděpodobnostním prostorem* nazveme trojici

$$(\Omega, \mathcal{A}, \mathbb{P}),$$

kde Ω je neprázdná množina, $\mathcal{A}$ je σ-algebra na Ω a $\mathbb{P}$ je pravděpodobnostní míra na $(\Omega, \mathcal{A})$. Prvkům $\mathcal{A}$ pak říkáme náhodné jevy.

Definice 13. *Náhodnou veličinou* na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathbb{P})$ nazveme takové zobrazení $X : \Omega \to \mathbb{R}$, že pro všechna $x \in \mathbb{R}$ platí:

$$\{\omega \in \Omega : \ X(\omega) < x\} \in \mathcal{A}.$$

Definice 14. *Distribuční funkcí* náhodné veličiny X na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathbb{P})$ nazveme funkci $F_X : \mathbb{R} \to \langle 0, 1\rangle$, jestliže pro všechna $x \in \mathbb{R}$ platí

$$F_X(x) = \mathbb{P}(\{\omega \in \Omega : \ X(\omega) \leq x\}).$$

Poznámka. Další zápisy, které budeme používat: $F_X(x) = \mathbb{P}_X((-\infty, x)) = \mathbb{P}(X \leq x)$.

Definice 15. *Diskrétní náhodnou veličinou* nazveme náhodnou veličinu, která nabývá konečně nebo spočetně mnoha hodnot.

Definice 16. *Spojitou náhodnou veličinou* nazveme náhodnou veličinu X s distribuční funkcí F_X, jestliže existuje taková nezáporná lesbeguevsky integrovatelná funkce $f_X : \mathbb{R} \to \langle 0, \infty)$, že pro všechna $x \in \mathbb{R}$ platí:

$$F_X(x) = \int_{-\infty}^{x} f_X(u) \ \mathrm{d}u.$$

Funkci f_X pak říkáme *hustota pravděpodobnostního rozdělení* náhodné veličiny X.

Poznámka. Hustotu pravděpodobnostního rozdělení někdy nazveme hustota pravděpodobnosti nebo zkráceně hustota.

Definice 17. *Střední hodnotou* diskrétní náhodné veličiny X nazveme číslo

$$E(X) = \sum_i x_i \mathbb{P}(X = x_i),$$

kde sčítáme přes všechny hodnoty x_i náhodné veličiny X.

Definice 18. *Střední hodnotou* spojité náhodné veličiny X s hustotou f_X

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) \, \mathrm{d}x.$$

Poznámka. Dále budeme místo $E(X)$ používat stručnější zápis EX.

Definice 19. Náhodné veličiny $X_1, X_2, \ldots, X_n$ na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathbb{P})$ nazveme *nezávislé*, jestliže pro libovolné intervaly $I_1, I_2, \ldots, I_n \subset \mathbb{R}$ platí:

$$\mathbb{P}(\bigcap_{i=1}^{n} \{\omega \in \Omega : \ X_i(\omega) \in I_i\}) = \prod_{i=1}^{n} \mathbb{P}(\{\omega \in \Omega : \ X_i(\omega) \in I_i\}).$$

Definice 20. *Mohutností množiny* rozumíme počet prvků z množiny A. Mohutnost množiny A značíme $\#A$.

Definice 21. *Diskrétní rovnoměrné rozdělení* je taková pravděpodobnostní míra $\mathbb{P}$ na $(\Omega, \mathcal{A})$, jestliže pro pravděpodobnost každé podmnožiny A množiny Ω platí

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}.$$

Poznámka. Z předchozí definice ihned plyne, že má-li $\mathbb{P}$ diskrétní rovnoměrné rozdělení na Ω, pak Ω má konečně mnoho prvků a $\mathcal{A}$ je systém všech podmnožin Ω.

Definice 22. *Spojité rovnoměrné rozdělení* je taková pravděpodobnostní míra $\lambda_{a,b}$ na $(\langle a, b), \mathcal{B}\langle a, b))$, že platí:

$$\lambda_{a,b}(\langle c, d\rangle) = \frac{d - c}{b - a},$$

jakmile $\langle c, d\rangle \subset \langle a, b)$.

Definice 23. Nechť $\omega \in \Omega$. *Diracovou mírou* nazveme pravděpodobnostní míru δ_ω na $(\Omega, \mathcal{A})$ takovou, že pro každé $A \in \mathcal{A}$

$$\delta_\omega(A) = \begin{cases} 1 & \text{pro } \omega \in A, \\ 0 & \text{pro } \omega \notin A. \end{cases}$$

Definice 24. Nechť je dána funkce $f : \Omega \to \mathbb{R}$ a $\mathbb{P}$ pravděpodobnostní míra na $(\Omega, \mathcal{A})$. Nechť $\sigma(f) \subset \mathcal{A}$, potom pravděpodobnostní míru $f_*\mathbb{P}$ na $(\mathbb{R}, \mathcal{B})$:

$$f_*\mathbb{P}(B) = \mathbb{P}(f^{-1}(B))$$

nazveme *indukovanou pravděpodobnostní mírou.*

Definice 25. Nechť jsou dány funkce $f : \mathbb{R} \to \mathbb{R}$ a $g : \mathbb{R} \to \mathbb{R}$ a nechť existuje $c \in \mathbb{R}^+$ a $x_0 \in \mathbb{R}$ takové, že platí $|f(x)| \leq c|g(x)|$, pro každé $x \geq x_0$. Potom píšeme

$$f(x) = O(g(x)).$$

Definice 26. *Charakteristickou funkcí* intervalu $\langle a, b) \subset \mathbb{R}$ nazveme funkci $\mathbb{I} : \mathbb{R} \to \{0, 1\}$, jestliže pro všechna $x \in \mathbb{R}$:

$$\mathbb{I}_{\langle a,b)}(x) = \begin{cases} 1 & \text{pro } x \in \langle a, b), \\ 0 & \text{pro } x \notin \langle a, b). \end{cases}$$

Část 1. **Benfordův zákon**

1.1. **Vznik.**

„To, že číslice z desítkové soustavy se neobjevují se stejnou frekvencí, musí být zřejmé každému, kdo často používá logaritmické tabulky a kdo si všimne, že první stránky jsou mnohem rychleji opotřebované než ty poslední". To jsou slova z článku [Ne] kanadského astronoma a matematika Simona Newcomba z roku 1881. Newcomb je námi první známý člověk, který zpozoroval, že frekvence, s jakou se signifikantní číslice vyskytují, se s většími číslicemi zmenšuje. Všiml si tak při listování logaritmických tabulek, kdy opotřebovanější první stránky mu vnukly myšlenku, že čísla začínající číslicí 1 jsou více vyhledávaná a používaná. Dále ve svém článku popisuje svůj objev jen po teoretické stránce. Jeho zjištění ho dovedlo k formulaci věty, která může být brána jako první definice rozdělení pravděpodobnosti související s Benfordovým zákonem. „Zákon pravděpodobnosti výskytu čísel je takový, že všechny mantisy jejich logaritmů jsou stejně pravděpodobné." Významem této vágně formulované věty si však nejsme příliš jisti. Pravděpodobně odpovídá Větě 60 ke které se dostaneme v části Charakterizace Benfordova zákona.

Poznámka. V našem případě je mantisa $\mathscr{M}_b$ chápána jako zlomková část logaritmu o zvoleném základu b, čísel větších než 1. Mantisy jsou si rovny pro součin čísel $1 \leq x < 10$ a $n-té$ mocniny b pro všechna $n \in \mathbb{N}$. Tedy platí $\mathscr{M}_b(x) = \mathscr{M}_b(x \cdot b^n)$. V případě výše citované věty Simona Newcomba, mantisy znamenaly pravděpodobně pouze zlomkovou část čísel.

V článku Simona Newcomba jsou i v tabulce vypsány pravděpodobnosti, s kterými se v číselném souboru vyhovujícímu tomuto zákonu vyskytují první a druhé možné číslice. Všímá si také, že pravděpodobnost dalších číslic v pořadí se čím dál více „přibližuje" k rovnoměrnému pravděpodobnostnímu rozdělení signifikantních číslic. Pravděpodobnosti, pro Newcombem vypsané první a druhé možné číslice, můžeme vidět v následující tabulce:

číslice	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
první číslice	–	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0458
druhá číslice	0.1197	0.1139	0.1088	0.1043	0.1003	0.0967	0.0934	0.0904	0.0876	0.0850

Tabulka 1. Pravděpodobnost výskytu první a druhé možné číslice dle Simona Newcomba.

Simon Newcomb po svém krátkém článku již nic k tématu Benfordova zákona nesepsal. Pravděpodobně díky tomu, že svůj článek nepodložil dalším zajímavým pozorováním, byl na dalších 57 let tento zákon zapomenut.

Elektroinženýr a fyzik Frank Benford, pak navázal tam kde Newcomb skončil a svoji studii popsal v [Ben]. Svůj výzkum podložil více jak 20 000 prvními číslicemi, sesbíranými z mnoha rozdílných zdrojů. Tyto číslice vykazovaly specifické logaritmické rozdělení pozorované Newcombem. Taktéž vypozoroval, že první číslice čísel z nesouvisejících témat, například všechna čísla ze stránek z novin, vyhovují tomuto logaritmickému rozdělení lépe, než zmiňované fyzikální a matematické konstanty.

1.2. **Signifikantní číslice.**
Po úvodu nám k úplnému popsání pravděpodobnosti s jakou se signifikantní číslice vyskytují v číselných souborech v případě, že vyhovují Benfordovu zákonu, schází jen definice samotných signifikantních číslic.

Definice 27. *První signifikantní číslicí* z desítkové soustavy nazveme zobrazení $D_1 : \mathbb{R} \to \{1, 2, \dots, 9\}$takové, že pro každé $x \in \mathbb{R} \setminus \{0\}$ platí: $\exists k \in \mathbb{Z}$:

$$10^k D_1(x) \le |x| < 10^k (D_1(x) + 1).$$

Poznámka. Pokud jsou ve studovaném číselném souboru nulové prvky, pak je obvykle z takového souboru vylučujeme.

Definice 28. Nechť $m \in \mathbb{N} \setminus \{1\}$ a je dáno zobrazení $D_m : \mathbb{R} \to \{0, 1, \dots, 9\}$. Pak řekneme, že D_m je m-tou signifikantní číslicí, jestliže pro každé $x \in \mathbb{R} \setminus \{0\}$ platí: $\exists k \in \mathbb{Z}$:

$$10^k \Big(\sum_{i=1}^{m-1} D_i 10^{m-1} + D_m(x) \Big) \le |x| < 10^k \Big(\sum_{i=1}^{m-1} D_i 10^{m-1} + D_m(x) + 1 \Big).$$

Poznámka. Signifikantní číslicí budeme v některých případech rozumět hodnotu zobrazení D_m. Zvolení symbolu D nebylo náhodné. Je převzato z anglického slova digit, neboli číslice. Všimněme si ještě rekurentního charakteru Definice 28:

Příklad 29. Mějme číslo -747.32, pak $D_1(-747.32) = 7$, protože pro $k = 2$ platí:

$$700 = 10^2 \cdot 7 \le |-747.32| < 10^2(7+1) = 800.$$

Dále $D_2(-747.32) = 4$, protože pro $k = 1$ platí:

$$740 = 10^1(7 \cdot 10 + 4) \le |-747.32| < 10^1(7 \cdot 10 + 4 + 1) = 750.$$

Příklad 30. Pro lepší představu o definici signifikantních číslic, z Definice 27 a Definice 28 Připomeňme, že $\sqrt{3} = 1.732\dots$, $\mathrm{e} = 2.718\dots$ a $\mathrm{e}^{-1} = 0.368\dots$.

$$D_1(\sqrt{3}) = D_1(-\sqrt{3}) = D_1(10\sqrt{3}) = 1,$$
$$D_2(\sqrt{3}) = 7,\ D_3(\sqrt{3}) = 3,$$
$$D_1(\mathrm{e}) = 2 \ne 3 = D_1(\mathrm{e}^{-1}).$$

1.3. **Benfordovo rozdělení signifikantních číslic.**
O Benfordově zákoně se hovoří jako o fenoménu první číslice, jelikož u prvních číslic je logaritmické rozdělení nejzřetelnější. Rovnice

$$\mathbb{P}(D_1 = d_1) = \log_{10}(1 + d_1^{-1}), \tag{1.1}$$

vyjadřuje pravděpodobnost, s jakou je signifikantní číslice D_1 rovna právě d_1, jestliže se číselný soubor „řídí“ Benfordovým zákonem. Už Newcomb ve svém článku [Ne] sleduje rozdělení pravděpodobnosti signifikantních číslic i u následujících signifikantních číslic v pořadí. Mějme tedy číselný soubor z desítkové soustavy který vyhovuje Benfordovu zákonu. Pak pravděpodobnost, s jakou se posloupnost m-signifikantních číslic $D_1, D_2, \ldots, D_m$, $m \in \mathbb{N}$, rovná právě posloupnosti čísel $d_1, d_2, \ldots, d_m$ v námi daném číselném souboru, je dána jako

$$\mathbb{P}((D_1, D_2, \ldots, D_m) = (d_1, d_2, \ldots, d_m)) = \log_{10}\left(1 + \left(\sum_{j=1}^{m} 10^{m-j} d_j\right)^{-1}\right). \tag{1.2}$$

Poznámka. Rovnice (1.2) lze také použít i k výpočtu pravděpodobnosti m-té signifikantní číslice. K tomu se dostaneme v Příkladu 32.

Z (1.2) je zřejmé, že žádný konečný číselný soubor nemůže úplně vyhovovat Benfordovu zákonu. Posloupnost čísel $d_1, d_2, \ldots, d_m$ lze totiž volit libovolně. Ideální číselný soubor vyhovující Benfordovu zákonu, musí proto být nutně nekonečný.

Příklad 31. Chtěli bychom tedy zjistit pravděpodobnost s jakou se vyskytnou číslice $1, 2, 3$ na prvních třech pozicích, ve stejném pořadí, v číselném souboru vyhovujícímu Benfordovu zákonu. Potom užitím (1.2)

$$\mathbb{P}((D_1, D_2, D_3) = (1, 2, 3)) = \log_{10}\left(1 + \frac{1}{1 \cdot 10^2 + 2 \cdot 10^1 + 3}\right) = \log_{10}\left(\frac{124}{123}\right) = 0.003517\ldots$$

V následujícím příkladu si ukážeme, že signifikantní číslice jsou na sobě závislé.

Příklad 32. Užitím vztahu (1.2) pro číselný soubor plně vyhovující Benfordovu zákonu. Pravděpodobnost, že druhá signifikantní číslice je rovna 2, je rovna součtu odpovídajících pravděpodobností $(D_1, 2)$.

$$\mathbb{P}(D_2 = 2) = \sum_{j=1}^{9} \mathbb{P}(D_1 = j, 2) = \sum_{j=1}^{9} \log_{10}\left(1 + \frac{1}{10j + 2}\right) = \log_{10}\left(\frac{13808461}{10747904}\right) = 0.108822\ldots$$

Podmíněná pravděpodobnost, že se druhá signifikantní číslice rovná 2 za předpokladu, že se první signifikantní číslice rovná 1, je:

$$\mathbb{P}(D_2 = 2 | D_1 = 1) = \frac{\mathbb{P}((D_1, D_2) = (1, 2))}{\mathbb{P}(D_1 = 1)} = \frac{\log_{10}\left(\frac{13}{12}\right)}{\log_{10} 2} = 0.115477\ldots$$

Poznámka. V případě nezávislosti signifikantních číslic by si byly pravděpodobnosti $P(D_2 = 2)$ a $P(D_2 = 2 | D_1 = 1)$ rovny.

Výše diskutovaná závislost signifikantních číslic se s rostoucím m exponenciálně zmenšuje. V [BH, str. 4, 5] se konstatuje, že podmíněnou pravděpodobnost m-té signifikantní číslice lze odhadnout takto:

$$\mathbb{P}(D_m = d_m | D_1 = d_1) = \mathbb{P}(D_m = d_m) + O(10^{-m}).$$

Podobně pravděpodobnost m-té signifikantní číslice se exponenciálně blíží k rovnoměrnému rozdělení signifikantních číslic. Pro $m \to \infty$ platí:

$$\mathbb{P}(D_m = d_m) = \frac{1}{10} + \frac{63}{20 \ln 10} 10^{-m} + O(10^{-2m}).$$

Poznámka. Symbol O viz Definice 25.

Po tomto úvodu do pravděpodobnostního rozdělení signifikantních číslic Benfordova zákona si už můžeme být jistější, že hovoříme o stejném zákoně, který měl na mysli Simon Newcomb.

Užitím (1.2), po zaokrouhlení, totiž získáme stejné hodnoty jako jsou v Tabulce 1 a také jsme si už potvrdili, že pravděpodobnostní rozdělení dalších signifikantních číslic v pořadí se dle Benfordova zákona blíží k rovnoměrnému rozdělení signifikantních číslic.

Ačkoliv tato vlastnost nebude v práci popsána, jelikož se budeme zabývat Benfordovým zákonem pro desítkovou soustavu, logaritmické rozdělení signifikantních číslic dle Benfordova zákona lze podobně nalézt ve všech celočíselných soustavách (se základem $b \geq 2$). Pravděpodobnost posloupnosti m signifikantních číslic (viz [BH, str. 5]) je dána jako:

$$\mathbb{P}((D_1^{(b)}, D_2^{(b)}, \ldots, D_m^{(b)}) = (d_1, d_2, \ldots, d_m)) = \log_b \left(1 + \left(\sum_{j=1}^{m} b^{m-j} d_j\right)^{-1}\right), \tag{1.3}$$

kde $\log_b$ je logaritmus o daném základě b. Signifikantní číslice $(D_1^{(b)}, D_2^{(b)}, \ldots, D_m^{(b)})$ o základu b mohou nabývat hodnot $d_1 \in \{1, 2, \ldots, b-1\}$ a pro $m \in \mathbb{N} \setminus \{1\}$ hodnoty $d_m \in \{0, 1, \ldots, b-1\}$.

Většina číselných souborů vyhovujících Benfordovu zákonu si tuto vlastnost zachovává i po transformaci do jiných číselných soustav. Celkově se tato vlastnost nazývá invariance ke změně základu. Jak si na jednom z příkladů ukážeme (viz Příklad 47), tato vlastnost neplatí vždy.

1.4. **Benfordovské číselné soubory.**

Jelikož Benfordův zákon vypovídá o pravděpodobnostním rozdělení signifikantních číslic v číselných souborech, můžeme si ukázat, jestli vůbec najdeme takové číselné soubory, které se tomuto logaritmickému rozdělení dle (1.2) „blíží“. Proto si ukážeme absolutní četnosti prvních signifikantních číslic dvou posloupností a porovnáme je s Benfordovým zákonem.

K lepšímu provnání shody využijeme maximální odchylky

$$R_{d_1} = \max_{d_1 \in \{1,\ldots,9\}} |\rho_N(d_1) - \log_{10}(1 + d_1^{-1})|,$$

kde $\rho_N(d_1) = \dfrac{\#\{n \in \mathbb{N} : 1 \leq n \leq N, \ D_1(x_n) = d_1\}}{N}$ je příslušná relativní četnost, (x_n) je zadaná posloupnost a $\log_{10}(1 + d_1^{-1})$ podle (1.1) vyjadřuje „očekávanou“ pravděpodobnost jevu $D_1(X) = d_1$, kde velké X je náhodná veličina, která se řídí dle Benfordova zákona. Tedy můžeme říci, že čím větší je R_{d_1}, tím menší shoda mezi dvěma číselnými soubory.

Příklad 33. Mějme číselný soubor C_N daný čísly z Fibonacciho posloupnosti (F_n). Tedy $S_N = \{f_1, f_2, \ldots, f_N\}$. Připomeňme rekurentní předpis, kterým je Fibonacciho posloupnost zadána. Nejprve položíme $f_1 = 1$, $f_2 = 1$. Další prvky Fibonacciho posloupnosti jsou pak součty dvou předchozích, tedy $f_3 = 1+1 = 2$, $f_4 = 1+2 = 3$, $f_5 = 2+3 = 5$ atd. V tabulce níže ke každé číslici d_1 přiřazujeme počet čísel z C_N s danou první signifikantní číslicí. Zkratka BZ v posledním řádku tabulky, odkazuje na „očekávanou“ pravděpodobnost náhodné veličiny řídící se Benfordovým zákonem.

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	R_{d_1} = [%]
$N = 10^2$	30	18	13	9	8	6	5	7	4	1.885
$N = 10^4$	3011	1761	1251	968	792	668	580	513	456	0.016
BZ = [%]	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*	–

Tabulka 2. Počet prvků C_N, začínajících danou signifikantní číslicí, pro N prvků.

Vidíme, že Fibonacciho posloupnost se pro sto prvních prvků shoduje s Benfordovým zákonem „docela dobře“. Pro $N = 10^4$ je tato shoda ještě větší.

Příklad 34. Ukažme si naopak, že existují zajímavé číselné soubory, které „nevyhovují“ Benfordovu zákonu. Mějme číselný soubor dán posloupností prvočísel (p_n). Odpovídající tabulka::

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$R_{d_1} = [\%]$
$N = 10^2$	25	19	19	20	8	2	4	2	1	10.309
$N = 10^4$	1601	1129	1097	1069	1055	1013	1027	1003	1006	14.093
$BZ = [\%]$	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*	–

TABULKA 3. Počet prvků začínajících danou signifikantní číslicí, pro prvních N prvků (p_n)

Pro signifikantní číslice prvních sto prvočísel nemusí být zřejmé, že se tato posloupnost čísel neřídí Benfordovým zákonem. Pro prvních 10^4 prvočísel se však i ta malá shoda dále zmenšuje. Na prvočíslech jsme si tedy ukázali, že existují číselné soubory, které nevykazují vlastnosti dané Benfordovým zákonem.

Podrobněji se danou problematikou budeme zabývat v Části VII.

Část 2. Funkce signifikandu a signifikandní σ-algebra

2.1. Funkce signifikandu.

Volně řečeno, podle [BH, str. 6] je signifikand čísla jeho koeficient, když je vyjádřeno v plovoucí řádové čárce. Funkce signifikandu nám v mnoha případech usnadní zápis pravděpodobnostního rozdělení signifikantních číslic. Upřesněme nyní pojem signifikandu.

Definice 35. *Funkcí signifikandu*, nazveme takovou funkci $S : \mathbb{R} \to \langle 1, 10)$, pro kterou platí následující tvrzení. Pro libovolné $x \in \mathbb{R} \setminus \{0\}$, existuje $t \in \langle 1, 10)$ takové, že platí:

$$S(x) = 10^{\log_{10}|x| - \lfloor \log_{10}|x| \rfloor} = t.$$

Navíc definujeme $S(0) = 0$.

Poznámka. Připomeňme, že $\lfloor x \rfloor$ označuje spodní celou část čísla x. Také si všimněme že $S(x) = S(10x)$. Ve většině případů signifikand bývá definován na intervalu $\langle 0, 1)$. V našem případě funkce signifikandu bude nabývat hodnot $\langle 1, 10)$, pro lepší zprostředkování výkladu Benfordova zákona.

Příklad 36. Zopakujme, že $\sqrt{3} = 1.732\ldots$, $\mathrm{e} = 2.718\ldots$ a $\mathrm{e}^{-1} = 0.368\ldots$. Z Definice 35 vyplývá

$$S(\sqrt{3}) = S(-\sqrt{3}) = S(10\sqrt{3}) = 1.732\ldots$$

$$S(\mathrm{e}) = 2.718\ldots, \; S(\mathrm{e}^{-1}) = 3.678\ldots$$

Funkce signifikandu také nijak neovlivňuje pořadí signifikantních číslic. Tedy jak Tvrzení 37 dole ukazuje, funkce signifikandu může být jednoznačně popsána pomocí signifikantních číslic a naopak.

Tvrzení 37. *Pro každé* $x \in \mathbb{R}$

(1) $S(x) = \sum_{m=1}^{\infty} 10^{1-m} D_m(x)$,
(2) $\forall m \in \mathbb{N} : D_m(x) = \lfloor 10^{m-1} S(x) \rfloor - 10 \lfloor 10^{m-2} S(x) \rfloor$.

Příklad 38. Z Tvrzení 37 vyplývá

$$S(\sqrt{3}) = D_1(\sqrt{3}) + 10^{-1} D_2(\sqrt{3}) + 10^{-2} D_3(\sqrt{3}) + 10^{-3} D_4(\sqrt{3}) \cdots = 1.732\ldots$$

$$D_1(\sqrt{3}) = \lfloor \sqrt{3} \rfloor = 1, \; D_2(\sqrt{3}) = \lfloor 10\sqrt{3} \rfloor - 10 \lfloor \sqrt{3} \rfloor = 7.$$

Ze vztahu mezi funkcí signifikandu a signifikantními číslicemi z Tvrzení 37 je zřejmé, že pravděpodobnost vzahující se k m-tici signifikantních číslic ze vztahu (1.2), lze přepsat pomocí funkce signifikandu. Tedy, pokud se náhodná veličina X řídí Benfordovým zákonem, pak pro všechna $1 \le t < 10$ platí

$$\mathbb{P}(S(X) \le t) = \log_{10} t. \tag{2.4}$$

Pro snadnější pochopení (2.4) si ukážeme tento rozdíl mezi zápisy odpovídajících pravděpodobnosti pomocí S a D.

Příklad 39. Pravděpodobnost, že „benfordovský soubor čísel“ bude mít první signifikantní číslici rovnou 2, je

$$\mathbb{P}(D_1 = 2) = \log_{10}(1 + \frac{1}{2}) = \log_{10}(\frac{3}{2}) = \log_{10} 3 - \log_{10} 2 = \mathbb{P}(2 \leq S < 3).$$

Poznámka. Vzhledem ke spojitosti funkce signifikandu S platí pravděpodobnost $\mathbb{P}(S \leq t)$=$\mathbb{P}(S < t)$, pro všechna $1 \leq t < 10$.

2.2. **Signifikandní σ-algebra.**

Z technického hlediska je nutné pro správné uchopení Benfordova zákona definovat vhodný pravděpodobnostní prostor. Budeme jej nazývat Benfordův pravděpodobnostní prostor. Pokud budeme hovořit o signifikantních číslicích nějakých čísel, pak na znaménku daného čísla nezáleží. Také číselné soubory, se kterými přirozeně pracujeme, jsou často složeny z kladných čísel. Proto první součástí Benfordova pravděpodobnostního prostoru je $\mathbb{R}^+$. Jako další určíme vhodnou σ-algebru.

Definice 40. Signifikandní *σ-algebrou $\mathscr{S}$* nazveme takovou σ-algebru na $\mathbb{R}^+$, která je generovaná funkcí signifikandu $S : \mathbb{R} \to \langle 1, 10)$, tzn.

$$\mathscr{S} = \mathbb{R}^+ \cap \sigma(S).$$

Signifikandní σ-algebra je tedy tvořena prvky $A \subset \mathbb{R}^+$, které mohou být popsány samotnou funkcí signifikandu. Pro libovolné $A \in \mathscr{S}$ a každé $x > 0$, funkce signifikandu dává postačující informaci k rozhodnutí jestli $x \in A$ nebo $x \notin A$. Podrobněji viz Lemma 41.

Lemma 41. *Buď dána funkce $f : \mathbb{R}^+ \to \mathbb{R}$. Potom pro každou funkci f jsou tato tvrzení ekvivalentní:*

(1) *f může být zcela popsána z hlediska funkce signifikandu S, tzn. pro vhodnou funkcí $\varphi : \langle 1, 10) \to \mathbb{R}$, pro kterou platí $\sigma(\varphi) \subset \mathcal{B}\langle 1, 10)$, platí:*

$$\forall x \in \mathbb{R}^+ : \ f(x) = \varphi(S(x))$$

(2) *$\sigma(f) \subset \mathscr{S}$.*

Důkaz.

(a) Nechť platí (1). Zvolme libovolný interval $I \subset \mathbb{R}$. Potom $B = \varphi^{-1}(I) \in \mathcal{B}\langle 1, 10)$ a $f^{-1}(I) = S^{-1}(\varphi^{-1}(I)) = S^{-1}(B)$, a jelikož $S^{-1}(B) \in \mathscr{S}$, pak $f^{-1}(I) \in \mathscr{S}$. Vzhledem k vlastnostem σ-algebry odtud plyne $\sigma(f) \subset \mathscr{S}$.

(b) Nechť platí (2). Nejdříve dokažme, že pro všechna $x > 0$ platí $f(10x) = f(x)$. Pro spor zvolme tedy například $x_0 \in \mathbb{R}^+$ tak, aby platilo $f(x_0) < f(10x_0)$. Mějme množinu $A \in \sigma(f) \subset \mathscr{S}$ takovou, že $x_0 \in A$ a naopak $10x_0 \notin A$. Protože $A = S^{-1}(B)$, pro vhodnou $B \in \mathcal{B}\langle 1, 10)$, pak $S(x_0) \in B$. Protože $S(x_0) = S(10x_0)$ a $S(10x_0) \notin B$ získáváme spor jelikož $S(x_0) \in B$ a zároveň $S(10x_0) \notin B$. Podobně lze vyvrátit opačnou nerovnost. Právě jsme dokázali, že pro všechna $x \in \mathbb{R}^+$ platí $f(10x) = f(x)$. Odtud snadno plyne, že pro každé $k \in \mathbb{Z}$ a každé $x > 0$ platí $f(10^k x) = f(x)$. Nyní pro $x \in \langle 1, 10)$ a jakékoliv $y > 0$ takové, že $S(y) = x$, definujme $\varphi(x) = f(y)$. Pak pro všechna y_i, y_j, kde $i, j \in \mathbb{N}$ taková, že $S(y_i) = S(y_j) = x$, $i \neq j$ liší o k-tou mocninu desítky, čili $y_i = 10^k y_j$, pro vhodné $k \in \mathbb{Z}$. Potom funkce $\varphi : \langle 1, 10) \to \mathbb{R}$ je korektně definovaná a $\varphi(S(y)) = f(y)$ platí pro všechna $y > 0$. Taktéž pro jakýkoliv interval $I \subset \mathbb{R}$ a $y > 0$, $\varphi(S(y)) \in I$ právě když, $y \in \bigcup_{k \in \mathbb{Z}} 10^k f^{-1}(I)$. Z (2) nyní plyne $\bigcup_{k \in \mathbb{Z}} 10^k f^{-1}(I) \in \mathscr{S}$, a proto $\sigma(\varphi) \subset \mathcal{B}\langle 1, 10)$. □

Připomeňme Tvrzení 37 ze kterého plyne, že každá množina, která lze popsat funkcí signifikandu, lze popsat signifikantními číslicemi. Z této vlastnosti plyne druhá část následující věty.

Věta 42. *Pro každé $A \in \mathscr{S}$, (připomeňme $S(A) = \{S(x) : x \in A\}$), platí*

$$A = \bigcup_{k \in \mathbb{Z}} 10^k S(A). \tag{2.5}$$

Mimoto,

$$\mathscr{S} = \mathbb{R}^+ \cap \sigma(D_1, D_2, D_3, \dots) = \{\bigcup_{k \in \mathbb{Z}} 10^k B : B \in \mathcal{B}\langle 1, 10)\}. \tag{2.6}$$

Důkaz. Z Definice 40 připomeňme vztah:

$$\mathscr{S} = \mathbb{R}^+ \cap \sigma(S) = \mathbb{R}^+ \cap \{S^{-1}(B) : B \in \mathcal{B}\langle 1, 10)\}.$$

Pro každé $A \in \mathscr{S}$ tedy existuje množina $B \in \mathcal{B}\langle 1, 10)$ taková, že $A = \mathbb{R}^+ \cap S^{-1}(B) = \bigcup_{k \in \mathbb{Z}} 10^k B$. Protože $S(A) = B$, pak platí (2.5) pro všechna $A \in \mathscr{S}$.

K důkazu 2.6 odkazujeme na [BH, Věta 2.9]. □

V následujícím příkladu pro lepší pochopení signifikandní σ-algebry popíšeme několik množin pro které zjistíme jestli jsou součástí $\mathscr{S}$.

Příklad 43. Množina A_1 kladných čísel, kterých první signifikantní číslice není rovna 2 a zároveň každá m-tá signifikantní číslice, $m \geq 2$, je buď 8, nebo 9, je prvkem signifikandní σ-algebry $\mathscr{S}$. Přesněji zapsáno

$$A_1 = \{x > 0 : D_1(x) \neq 2, D_m(x) \in \{8, 9\}, m \geq 2\} \in \mathscr{S}.$$

Dále

$$A_2 = \{x > 0 : 3 \leq S(x) < 4)\} \in \mathscr{S},$$

$$A_3 = \langle 3, 4) \notin \mathscr{S}.$$

Poznámka. Množina A_3 nevyhovuje (2.5), jelikož kladná čísla $x \in \ldots \cup \langle 0.3, 0.4) \cup \langle 30, 40) \cup \langle 300, 400) \cup \ldots$ nejsou součástí této množiny.

Množina

$$A_4 = \{10^k + 2 \cdot 10^{k-1} : k \in \mathbb{Z}\} = \{\dots, 0.012, 0.12, 1.2, 12, 120, \dots\}$$

je prvkem signifikandní σ-algebry. Pro snadnější pochopení, množinu A_4 můžeme přepsat jako množinu kladných čísel, kterých první signifikantní číslice je rovna 1 a druhá signifikantní číslice je rovna 2, čili $A_4 = \{x > 0 : D_1(x) = 1, D_2(x) = 2\}$.

Lemma 44. *Pro signifikandní σ-algebru $\mathscr{S}$ platí:*

(1) *Pro každé $A \in \mathscr{S}$ a $k \in \mathbb{Z}$:*

$$10^k A = A.$$

(2) *Signifikandní σ-algebra je uzavřená vůči násobení skalárem. To znamená, že pro každé $A \in \mathscr{S}$ a $\alpha > 0$,*

$$\alpha A \in \mathscr{S}.$$

(3) *Signifikandní σ-algebra je uzavřená vůči mocnině převráceného přirozeného čísla. Pro každé $A \in \mathscr{S}$ a $n \in \mathbb{N}$:*

$$A^{1/n} \in \mathscr{S}.$$

Poznámka. Upřesněme, že výše uvedené algebraické operace na množinách chápeme ve významu jejich aplikace na všechny prvky daných množin.

Poznámka. $\mathscr{S}$ není uzavřená vůči všem mocninám přirozeného čísla. Pro $A \in \mathscr{S}$ je $A^n \in \mathscr{S}$ právě tehdy, jestliže pro některou $B \in \mathcal{B}\langle 1, 10)$ platí:

$$S(A)^n = B \cup 10B \cup \cdots \cup 10^{n-1}B. \tag{2.7}$$

Důkaz.

(1) Důkaz je zřejmý z (2.5), jelikož $S(10^k A) = S(A)$, pro každé $k \in \mathbb{Z}$.

(2) Z (2.6) víme, že pro $A \in \mathscr{S}$ existuje $B \in \mathcal{B}\langle 1, 10)$ takové, že $A = \bigcup_{k\in\mathbb{Z}} 10^k B$. Bez újmy na obecnosti předpokládejme $1 < \alpha < 10$. Pak

$$\alpha A = \bigcup_{k\in\mathbb{Z}} 10^k \alpha B = \bigcup_{k\in\mathbb{Z}} 10^k \Big((\alpha B \cap \langle \alpha, 10)) \cup (\frac{\alpha}{10} B \cap \langle 1, \alpha)) \Big) = \bigcup_{k\in\mathbb{Z}} 10^k C,$$

čili $C = (\alpha B \cap \langle \alpha, 10)) \cup (\frac{\alpha}{10} B \cap \langle 1, \alpha)) \in \mathcal{B}\langle 1, 10)$, z čehož vyplývá $\alpha A \in \mathscr{S}$.

(3) Pro aplikaci (2.6) nejprve zvažme, že pro všechna $0 \le r < s < 1$ platí $\langle 10^r, 10^s) \in \mathcal{B}\langle 1, 10)$ a protože $\mathcal{B}\langle 1, 10)$ je intervaly daného typu generována, dokážeme platnost (3) pro speciální případ $A = \bigcup_{k\in\mathbb{Z}} 10^k \langle 10^r, 10^s)$. V tomto případě

$$A^{1/n} = \bigcup_{k\in\mathbb{Z}} 10^{k/n} \langle 10^{r/n}, 10^{s/n}) = \bigcup_{k\in\mathbb{Z}} 10^k \bigcup_{j=0}^{n-1} \langle 10^{(r+j)/n}, 10^{(s+j)/n}) = \bigcup_{k\in\mathbb{Z}} 10^k C,$$

kde $C = \bigcup_{j=0}^{n-1} \langle 10^{(r+j)/n}, 10^{(s+j)/n}) \in \mathcal{B}\langle 1, 10)$, z čehož vyplývá $A^{1/n} \in \mathscr{S}$.

□

Příklad 45. Nechť

$$A_5 = \{x > 0 : \ 1 \le S(x) < 2)\}.$$

Zřejmě A_5 je prvkem $\mathscr{S}$. Z Věty 42 připomeňme (2.5) a (2.6). Tedy

$$A_5 = \{x > 0 : \ D_1(x) = 1\} = \bigcup_{k\in\mathbb{Z}} 10^k \langle 1, 2).$$

K ilustraci bodu (2) Lemmatu 44, zvolme $\alpha = 7$. Potom

$$\begin{aligned} 7A_5 &= \bigcup_{k\in\mathbb{Z}} 10^k \langle 7, 14) = \{x > 0 : \ S(x) \in \langle 1, 1.4) \cup \langle 7, 10)\} = \\ &= \{x > 0 : \ (D_1(x) \in \{7, 8, 9\}) \cup (D_1(x) = 1 \cap D_2(x) \in \{0, 1, 2, 3\})\} \in \mathscr{S}. \end{aligned}$$

Nelze si nevšimnout, že při specifikaci množiny signifikantními číslicemi, by v některých případech byl zápis zbytečně komplikovaný. V dalších takových případech proto od tohoto zápisu upustíme.

K Lemmatu 44 (3). Pro $n = 2$ získáváme množinu

$$A_5^{1/2} = \bigcup_{k\in\mathbb{Z}} 10^k (\langle 1, \sqrt{2}) \cup \langle \sqrt{10}, \sqrt{20})) = \{\{x > 0 : \ S(x) \in \langle 1, \sqrt{2}) \cup \langle \sqrt{10}, \sqrt{20})\}\} \in \mathscr{S}.$$

Ilustrujme ještě poznámku k Lemmatu 44. Zvolme $n = 2$. Pak

$$A_5^2 = \bigcup_{k\in\mathbb{Z}} 10^{2k}\langle 1,4) \notin \mathscr{S},$$

(protože intervaly $\bigcup_{k\in\mathbb{Z}} 10^{2k-1}\langle 1,4)$ nejsou součástí množiny A_5^2, potom A_5^2 není množinou signifikandní σ-algebry).

Uvažujme ještě jednou množinu

$$A_5^{1/2} = \bigcup_{k\in\mathbb{Z}} 10^k \left(\langle 1,\sqrt{2}) \cup \langle\sqrt{10},\sqrt{20})\right).$$

Pak

$$(A_5^{1/2})^2 = \bigcup_{k\in\mathbb{Z}} 10^{2k} \left(\langle 1,2) \cup \langle 10,20)\right) \in \mathscr{S}.$$

Výše uvedené tvrzení je zřejmé, za povšimnutí však stojí zápis, ze kterého lze snadno pochopit význam (2.7).

Část 3. Benfordovské matematické struktury

Pravděpodobnostní rozdělení v (1.2), (2.4), byla zatím pro jednoduchost definovány pouze vůči „číselným souborům“. Abychom lépe porozuměli Benfordovu zákonu, je nutné si toto logaritmické rozdělení signifikantních číslic definovat i vůči některým matematickým strukturám. K tomuto účelu byly vybrány reálné posloupnosti, funkce kladných reálných čísel, pravděpodobnostní rozdělení a náhodné veličiny.

Vzhledem ke složitosti některých výpočtů v následujících příkladech pouze naznačíme, které dané matematické struktury jsou benfordovské. Intuitivně zvolenými kroky není však příliš těžké zjistit, že některé dané matematické struktury benfordovské nejsou. V následující části práce si některé tyto příklady řádně vyřešíme.

3.1. Benfordovské posloupnosti.

Definice 46. *Benfordovskou posloupností* nazveme posloupnost čísel $(x_n) = \{x_n : x_n \in \mathbb{R}, n \in \mathbb{N}\}$, pokud pro všechna $t \in \langle 1, 10)$:

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : S(x_n) \le t\}}{N} = \log_{10} t, \tag{3.8}$$

nebo ekvivalentně: $\forall m \in \mathbb{N}$, $\forall d_1 \in \{1, 2, \dots, 9\}$, $\forall d_j \in \{0, 1, 2, \dots, 9\}$, $2 \le j \in \mathbb{N}$:

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : D_i(x_n) = d_i, i = 1, 2, \dots, m\}}{N} = \log_{10} \Big(1 + (\sum_{j=1}^{m} 10^{m-i} d_i)^{-1}\Big). \tag{3.9}$$

Poznámka. $\#\{1 \le n \le N : D_i(x_n) = d_i, \forall i \in \mathbb{N}\}$ označuje počet prvků posloupnosti (x_n), vyhovujících umístěním požadovaných signifikantních číslic.

Jinak řečeno, posloupnost reálných čísel (x_n), $n \in \mathbb{N}$ je benfordovská, pokud se relativní četnosti signifikantních číslic na (x_n) blíží k pravděpodobnostnímu rozdělení dle (1.2) a (2.4).

Příklad 47. Jak zjistíme později, posloupnost $(x_n) = (2^n)$ je benfordovská. V následující tabulce vidíme, že shoda posloupnosti (2^n) s číselným souborem vyhovujícím Benfordovu zákonu je ještě vyšší než u Fibonacciho čísel v Tabulce 2.

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$R_d = [\%]$
$n = 10^2$	30	17	13	10	7	7	6	5	5	0.918
$n = 10^4$	3010	1761	1249	970	791	670	579	512	458	0.009
$BZ = [\%]$	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*	–

Tabulka 4. Počet prvků (2^n), začínajících danou signifikantní číslicí, pro n prvků.

U posloupnosti (2^n) nastává ten případ, kdy víme, že není benfordovská v jiné číselné soustavě, a to ve dvojkové. Posloupnost (2^n) má ve dvojkové soustavě každou signifikantní číslici kromě první rovnou nule. Tedy nevyhovuje (1.3).

Příklad 48. Posloupnost $(x_n) = (n)$, $n \in \mathbb{N}$ benfordovská není. Uveďme zatím intuitivní argument. Rozdělením přirozených čísel do množin tvaru $10^i\langle 1, 10) \cap \mathbb{N}$, $(i \in \mathbb{N})$ zjistíme, že rozdělení

$(D_1, D_2, \ldots, D_{i+1})$ signifikantních číslic je na každé z uvažovaných množin „rovnoměrné".

3.2. **Benfordovské funkce.**

Definice 49. *Benfordovskou funkcí* nazveme borelovsky měřitelnou funkci $f : \langle 0, +\infty) \to \mathbb{R}$ takovou, že pro všechna $t \in \langle 1, 10)$ platí:

$$\lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0, T) : S(f(\tau)) \leq t\})}{T} = \log_{10} t. \tag{3.10}$$

Ekvivalentně: $\forall m \in \mathbb{N}$, $\forall d_1 \in \{1, 2, \ldots, 9\}$, $\forall d_j \in \{0, 1, 2, \ldots, 9\}$, $2 \leq j \in \mathbb{N}$:

$$\lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0, T) : D_i(f(\tau)) = d_i, i = 1, 2, \ldots, m\})}{T} = \log_{10} \Big(1 + (\sum_{i=1}^{m} 10^{m-i} d_i)^{-1}\Big). \tag{3.11}$$

Poznámka. $\lambda(\{\tau \in \langle 0, T) : D_i(f(\tau)) = d_i, i = 1, 2, \ldots, m\})$ (viz Definice 9) vyjadřuje „délku množiny" všech $\tau \in \langle 0, T)$, pro které $f(\tau)$ vyhovuje umístěním požadovaných signifikantních číslic.

Obdobně jako u benfordovských posloupností, tedy funkci nazveme benfordovskou, jestliže se rozdělení signifikantních číslic funkčních hodnot $f(\tau)$ blíží pravděpodobnostnímu rozdělení vyhovujícímu Benfordovu zákonu pro $\tau \in \langle 0, T)$ s $T \to \infty$.

Příklad 50. Později si ukážeme, že $f(t) = \mathrm{e}^{\alpha t}$ je benfordovská funkce pro všechna $\alpha \neq 0$.

Příklad 51. Podobně jako u posloupnosti přirozených čísel můžeme o funkci $f(t) = t$ pro $t \in \mathbb{R}^+$ říct, že není benfordovská, protože „na všech intervalech $10^k\langle 1, 10)$, pro $k \in \mathbb{Z}$, jsou všechny m-té signifikantní číslice rovnoměrně rozdělené".

Příklad 52. Roli funkce f z Definice 49 nechť zaujme funkce sinus. Pro vhodný spodní odhad levé strany v (3.10) volíme $|\sin(t)| \in \langle 0.9, 1)$. To znamená, že vezmeme některé funkční hodnoty, jejichž první signifikantní číslice je 9. Potom víme, že na intervalu $\langle 0, \frac{\pi}{2})$ je „příslušná pravděpodobnost" volby funkční hodnoty z $\langle 0.9, 1)$ dána jako:

$$\rho_1(|f(t)| \in \langle 0.9, 1)) = \frac{(\dfrac{\pi}{2} - \arcsin(0.9))}{\frac{\pi}{2}} = 0.287133\ldots.$$

Ze symetrie funkce sinus plyne, že pravděpodobnost $\rho_2(|f(t)| \in \langle 0.9, 1))$ odpovídající intervalu $(\frac{\pi}{2}, \pi\rangle$ je rovna $\rho_1(|f(t)| \in \langle 0.9, 1))$. Opět ze symetrie plyne rovnost $\rho_1(|f(t)| \in \langle 0.9, 1)) = \rho_3(|f(t)| \in \langle 0.9, 1))$, kde $\rho_3(|f(t)| \in \langle 0.9, 1))$ odpovídá pravděpodobnosti na intervalu $(\pi, 2\pi\rangle$. Vzhledem k libovolnému intervalu délky 2π, tedy platí

$$\rho(|f(t)| \in \langle 0.9, 1)) = \rho_1(|f(t)| \in \langle 0.9, 1)).$$

Jelikož je funkce sinus 2π periodická, limitním přechodem zjistíme, že

$$\lim_{k\to\infty} \frac{\lambda(\{\tau \in \langle 0, 2k\pi) : 9 \leq S(\sin(\tau)) < 10\})}{2k\pi} \geq \rho_1(|f(t)| \in \langle 0.9, 1)),$$

a jelikož je $\rho_1(|f(t)| \in \langle 0.9, 1)) > \log_{10}(\frac{10}{9})$, rozdělení signifikantních číslic ve funkci sinus nevyhovuje logaritmickému rozdělení dle (1.1). Tedy ani nevyhovuje (3.10) a (3.11).

3.3. **Benfordovské rozdělení a náhodné veličiny.**

Definice 53. *Benfordovskou pravděpodobnostní mírou* nazveme pravděpodobnostní míru P na $(\mathbb{R}, \mathcal{B})$, pokud pro všechna $t \in \langle 1, 10)$ platí:

$$P(\{x \in \mathbb{R} : \ S(x) \le t\}) = \log_{10} t. \tag{3.12}$$

Benfordovskou náhodnou veličinou nazveme náhodnou veličinu X na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathbb{P})$, jestliže je pravděpodobnostní míra P_X *benfordovskou.* Tedy

$$\mathbb{P}(S(X) \le t) = P_X(\{x \in \mathbb{R} : \ S(x) \le t\}) = \log_{10} t$$

nebo pro všechna $m \in \mathbb{N}$, $\forall d_1 \in \{1, 2, \ldots, 9\}$, $\forall d_j \in \{0, 1, 2, \ldots, 9\}$, $2 \le j \in \mathbb{N}$

$$\mathbb{P}(D_i(X) = d_i, i = 1, 2, \ldots, m) = \log_{10} \Big(1 + (\sum_{i=1}^{m} 10^{m-i} d_i)^{-1}\Big).$$

Poznámka. Poznamenejme, že benfordovskou pravděpodobnostní míru lze také zapsat jako $S_* P(\{0\} \cup \langle 1, t\rangle) = \log_{10} t$.

Příklad 54. Pro každé $k \in \mathbb{Z}$ je pravděpodobnostní míra P_k na $\langle 10^k, 10^{k+1})$ s hustotou $f_k(x) = \frac{1}{x \ln 10}$ benfordovská. To je zřejmé z (3.12). Také $\frac{1}{2}(P_k + P_{k+1})$ je benfordovská pravděpodobnostní míra, protože je to součet polovičních hodnot stejného pravděpodobnostního rozdělení. Pak tedy existují i $q_k \in \langle 0, 1\rangle$ takové, pro které $\sum_{k=\mathbb{Z}} q_k P_k$ je benfordovská pravděpodobnostní míra, když $\sum_{k \in \mathbb{Z}} q_k = 1$.

Příklad 55. Mějme náhodnou veličinu X na $(\langle 0, 1), \mathcal{B}\langle 0, 1), \lambda_{0,1})$, kde $\lambda_{0,1}$ je spojité rovnoměrné pravděpodobnostní rozdělení (viz Definice 22) na intervalu $\langle 0, 1)$. Jelikož jsou všechny zkoumané intervaly s požadovanými signifikantními číslicemi na intervalech $10^{-n}\langle 1, t\rangle$, $n \in \mathbb{N}$, pak pro všechna $1 \le t < 10$ můžeme psát

$$\mathbb{P}(S(X) \le t) = \lambda_{0,1}(\bigcup_{n=1}^{\infty} 10^{-n}\langle 1, t\rangle) = \sum_{n=1}^{\infty} 10^{-n}(t-1) = \frac{t-1}{9} \not\equiv \log_{10} t.$$

Tedy X není benfordovská náhodná veličina.

Příklad 56. Mějme náhodnou veličinu X danou exponenciálním rozdělením se střední hodnotou 1, tzn. $X \to E(1)$. Distribuční funkce X je dána jako $F_X(t) = 1 - e^{-t}$, $t > 0$. Potom

$$\mathbb{P}(D_1(X) = 2) = \mathbb{P}(X \in \bigcup_{k \in \mathbb{Z}} 10^k \langle 2, 3)) = \sum_{k \in \mathbb{Z}} (e^{-10^k \cdot 2} - e^{-10^k \cdot 3}) \approx 17.4322.$$

Jelikož pravděpodobnost daná Benfordovým zákonem pro první signifikantní číslici rovnou 2 je $\log_{10} \frac{3}{2} = 0.17609\ldots$, pak exponenciální rozdělění se střední hodnotou 1 není benfordovské.

3.4. **Benfordovo pravděpodobnostní rozdělení.**

Teď na chvíli upustíme od samotného poznávání benfordovských matematických struktur a nadefinujeme si poslední součást Benfordova pravděpodobnostního prostoru.

Definice 57. *Benfordovo pravděpodobnostní rozdělení* je specifická pravděpodobnostní míra $\mathbb{B}$ na $(\mathbb{R}^+, \mathscr{S})$, pokud pro všechna $t \in \langle 1, 10)$

$$\mathbb{B}(S \le t) = \mathbb{B}(\bigcup_{k \in \mathbb{Z}} 10^k \langle 1, t\rangle) = \log_{10} t \tag{3.13}$$

nebo pro všechna $m \in \mathbb{N}$, $\forall d_1 \in \{1, 2, \ldots, 9\}$, $\forall d_j \in \{0, 1, 2, \ldots, 9\}$, $2 \leq j \in \mathbb{N}$

$$\mathbb{B}(D_i = d_i, i = 1, 2, \ldots, m) = \log_{10}\Big(1 + (\sum_{i=1}^{m} 10^{m-i} d_i)^{-1}\Big).$$

Je dobré mít na paměti, že $\mathbb{B}$ je definováno vzhledem k signifikandní σ-algebře, funkci signifikandu a signifikantním číslicím, a bude takto používáno. Pravděpodobnostní rozdělení $\mathbb{B}$ nám v následujících částech umožní lépe charakterizovat Benfordův zákon a rozeznat benfordovské matematické struktury.

Následující Lemma vypovídá o převodu mezi pravděpodobnostní mírou $\mathbb{P}$ na $(\mathbb{R}^+, \mathscr{S})$ a pravděpodobnostní mírou P na $(\langle 0,1), \mathcal{B}\langle 0,1))$. Pravděpodobnostní prostor $(\langle 0,1), \mathcal{B}\langle 0,1), P)$ označujeme jako klasický. Tento převod nám v následující části práce pomůže lépe charakterizovat Benfordův zákon.

Lemma 58. *Funkce $\ell : \mathbb{R}^+ \to \langle 0,1)$ definovaná jako $\ell(x) = \log_{10}(S(x))$ jednoznačně stanovuje izomorfismus míry mezi*

(1) *pravděpodobnostní mírou $\mathbb{P}$ na $(\mathbb{R}^+, \mathscr{S})$*
(2) *pravděpodobnostní mírou P na $(\langle 0,1), \mathcal{B}\langle 0,1))$.*

Důkaz. Odkazujeme [BH, Lemma 2.6]. □

Jak se později ukáže, volba funkce ℓ z Lemmatu 58 nebyla náhodná. Společně s $\mathbb{B}$ totiž indukují rovnoměrnou pravděpodobnostní míru $\lambda_{0,1}$ na $(\langle 0,1), \mathcal{B}\langle 0,1))$.

Část 4. Charakterizace Benfordova zákona

Tato část práce nám umožní lépe charakterizovat benfordovské matematické struktury a s tím i benfordovské pravděpodobnostní rozdělení skrze další zajímavé vlastnosti Benfordova zákona.

4.1. Charakterizace rovnoměrným rozdělením.

Rovnoměrné rozdělení je pravděpodobně nejintuitivnější rozdělení nám známé. Právě rovnoměrné rozdělení modulo 1 má mnoho vlastností, s kterými snadněji poznáme benfordovské matematické struktury. Matematické struktury, které v následujícím textu budeme studovat, jsou znovu reálné posloupnosti, funkce kladných reálných čísel, pravděpodobnostní rozdělení a náhodné veličiny.

Ke správné definici matematických struktur, které se řídí rovnoměrným rozdělením modulo 1, si funkcí $\mathcal{F} : \mathbb{R} \to \langle 0, 1)$ označíme zlomkovou část. Pak tedy pro všechna $x \in \mathbb{R}$ platí

$$\mathcal{F}(x) = x - \lfloor x \rfloor.$$

Příklad. $\mathcal{F}(\pi) = 0.1415\dots$, $\mathcal{F}(-2.375) = -2.375 + 3 = 0.625$.

Definice 59. Posloupností *rovnoměrně rozdělenou modulo* 1 nazveme takovou posloupnost (x_n), že pro všechna $s \in \langle 0, 1)$

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : \mathcal{F}(x_n) \le s\}}{N} = s. \tag{4.14}$$

Funkcí *rovnoměrně rozdělenou modulo* 1 nazveme takovou borelovsky měřitelnou funkci $f : \langle 0, +\infty) \to \mathbb{R}$, že pro všechna $s \in \langle 0, 1)$

$$\lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0, T) : \mathcal{F}(f(\tau)) \le s\})}{T} = s. \tag{4.15}$$

Pravděpodobnostní mírou *rovnoměrně rozdělenou modulo* 1 nazveme takovou pravděpodobnostní míru P na $(\mathbb{R}, \mathcal{B})$, že pro všechna $s \in \langle 0, 1)$

$$P(\{x : \mathcal{F}(x) \le s\}) = P(\bigcup_{k\in\mathbb{Z}} \langle k, k + s\rangle) = s.$$

Náhodnou veličinu *rovnoměrně rozdělenou modulo* 1 nazveme takovou náhodnou veličinu X na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathbb{P})$, že pro všechna $s \in \langle 0, 1)$

$$\mathbb{P}(\mathcal{F}(X) \le s) = s.$$

Poznámka. Existuje spojitost mezi rovnoměrným rozdělením modulo 1 a rovnoměrným rozdělením na $\langle 0, 1)$. Máme-li například zmiňovanou rovnoměrně rozdělenou náhodnou veličinu modulo 1, pak $\mathbb{P}(\mathcal{F}(X) \le s) = \lambda_{0,1}(\langle 0, s\rangle)$.

Následující věta nám nabídne podstatné pojítko mezi Benfordovým zákonem a rovnoměrným pravděpodobnostním rozdělením modulo 1. Vzhledem k tomu, že budeme užívat logaritmy matematických struktur, které můžou nabývat hodnoty 0, definujme $\log_b 0 = 0$.

Věta 60. *Posloupnost reálných čísel, borelovsky měřitelná funkce, náhodná veličina a borelovská pravděpodobnostní míra jsou benfordovské právě tehdy, když dekadický logaritmus jejich absolutní hodnoty je rovnoměrně rozdělený modulo* 1.

Důkaz. Buď dána funkce $f : \mathbb{R} \to \mathbb{R}$, pak pro všechna $s \in \langle 0, 1)$,

$$\lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0,T) : \mathcal{F}(\log_{10}|f(\tau)|) \le s\})}{T} = \lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0,T) : \log_{10}|f(\tau)| \in \bigcup_{k\in\mathbb{Z}}\langle k, k+s)\})}{T} =$$

$$= \lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0,T) : |f(\tau)| \in \bigcup_{k\in\mathbb{Z}} 10^k\langle 1, 10^s)\})}{T} = \lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0,T) : S(f(\tau)) \le 10^s\})}{T}.$$

Užitím Definice 49 a Definice 59 je funkce f je benfordovská právě tehdy, když

$$\lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0,T) : S(f(\tau)) \le 10^s\})}{T} = \log_{10} 10^s = s.$$

Stejnými kroky lze toto dokázat i pro ostatní matematické struktury.

Další Lemma nám opět rozšíří pole vlastností týkajících se rovnoměrného rozdělení modulo 1 a tedy i Benfordova zákona. □

Lemma 61.

(1) *Jestliže* $\lim_{n\to\infty} |y_n - x_n| = 0$. *Potom posloupnost čísel* (x_n) *je rovnoměrně rozdělená modulo* 1 *právě tehdy, když posloupnost* (y_n) *je rovnoměrně rozdělená modulo* 1.

(2) *Posloupnost čísel* (x_n) *je rovnoměrně rozdělená modulo* 1 *právě tehdy, když posloupnost* $(kx_n + b)$ *je rovnoměrně rozdělená modulo* 1 *pro všechna* $k \in \mathbb{Z} \setminus \{0\}$ *a všechna* $b \in \mathbb{R}$.

(3) *Funkce* f *je rovnoměrně rozdělená modulo* 1 *právě tehdy, když* $t \to kf(t)+b$ *je rovnoměrně rozdělené modulo* 1 *pro všechna* $k \in \mathbb{Z} \setminus \{0\}$ *a všechna* $b \in \mathbb{R}$.

(4) *Náhodná veličina* X *je rovnoměrně rozdělená modulo* 1 *právě tehdy, když* $kX + b$ *je rovnoměrně rozdělená modulo* 1 *pro všechna* $k \in \mathbb{Z} \setminus \{0\}$ *a všechna* $b \in \mathbb{R}$.

Důkaz. Odkazujeme na [BH, Lemma 4.3]. □

V následujících příkladech budeme zmiňovat hlavně ty případy, které nám v kombinaci s Lemmatem 60 potvrdí nebo vyvrátí benfordovskost matematické struktury.

Příklad 62. Vzpomeňme si na Příklad 48 s posloupností (n). Užitím Lemmatu 60 dostaneme:

$$\liminf_{N\to\infty} \frac{\#\{1 \le n \le N : \mathcal{F}(\log_{10} n) \le s\}}{N} == \frac{1}{9}(10^s - 1),$$

$$\limsup_{N\to\infty} \frac{\#\{1 \le n \le N : \mathcal{F}(\log_{10} n) \le s\}}{N} = \frac{10}{9}(1 - 10^{-s}).$$

Jelikož si lim inf a lim sup nejsou rovny, potom n nemůže být benfordovská posloupnost jelikož $\log_{10} n$ není rovnoměrně rozdělená modulo 1.

Příklad 63. Mějme funkci $f(t) = at + b$ pro všechna reálná $a \ne 0$ a b. Pro $a = 0$ je funkce konstantní, čili nemůže být rovnoměrně rozdělená modulo 1. Mějme tedy $a > 0$, pak všechna $t \in \mathbb{R}$, pro $b \in \mathbb{R}$ takové, že $\mathcal{F}(at + b) \le s$, platí že $t \in \bigcup_{k\in\mathbb{Z}}\langle \frac{k-b}{a}, \frac{k-b+s}{a}\rangle$. Po vzoru (4.15) určíme čitatel $\lambda(\{t \in \langle 0,T) : \mathcal{F}(at + b) \le s\})$. Pro $t \in \langle 0,T)$ takové, že $\mathcal{F}(at + b) \le s$

$$\frac{k-b}{a} \ge 0 \implies k \ge \lfloor b \rfloor,$$

$$\frac{k-b+s}{a} \le T \implies k \le \lceil aT + b - s \rceil$$

Všimněme si, že $\lambda(\langle\frac{k-b}{a},\frac{k-b+s}{a}\rangle)=\frac{s}{a}$, proto $\lambda(\{t\in\langle 0,T)\,:\,\mathcal{F}(at+b)\le s\})$ je rovna

$$\lambda(\bigcup_{k=\lfloor b\rfloor}^{\lceil aT+b-s\rceil}\langle\frac{k-b}{a},\frac{k-b+s}{a}\rangle)=\begin{cases}\frac{s}{a}(\lceil aT\rceil+1) & \text{pro } \mathcal{F}(aT)+\mathcal{F}(b)-s\ge 1\\ \frac{s}{a}\lceil aT\rceil & \text{pro } 0\le\mathcal{F}(aT)+\mathcal{F}(b)-s<1\\ \frac{s}{a}(\lceil aT\rceil-1) & \text{pro } \mathcal{F}(aT)+\mathcal{F}(b)-s<0\end{cases}.$$

Tedy dle (4.15) $\lim_{T\to\infty}\frac{s}{aT}(\lceil aT\rceil+1)=s$, dále $\lim_{T\to\infty}\frac{s}{aT}(\lceil aT\rceil)=s$ a $\lim_{T\to\infty}\frac{s}{aT}(\lceil aT\rceil-1)=s$, pak $f(t)=at+b$ je rovnoměrně rozdělená funkce modulo 1 pro všechna $a>0$ a všechna reálná b. Stejnými kroky lze toto zjistit i pro $a<0$.

To nás přivádí k Příkladu 50 a k funkci $f(t)=\mathrm{e}^{\alpha t}$. Potom užitím Lemmatu 60 platí, $\log_{10}f(t)=\frac{\alpha}{ln10}t$, funkce $\log_{10}f(t)$ je proto rovnoměrně rozdělená modulo 1 a funkce $f(t)$ je benfordovská.

Důsledek 64. Odvozeno z Definice 59, Věty 60 a Lemmatu 61.

(1) *Posloupnost čísel (x_n) je benfordovská právě tehdy, když pro všechna $\alpha\in\mathbb{R}$ a $\forall k\in\mathbb{Z}$, kde $\alpha k\neq 0$, je posloupnost (αx_n^k) benfordovská.*

(2) *Funkce $f:\langle 0,\infty)\to\mathbb{R}$ je benfordovská právě tehdy, když $\frac{1}{f}$ je benfordovská funkce.*

(3) *Náhodná veličina X je benfordovská právě tehdy, když $\frac{1}{X}$ je benfordovská náhodná veličina.*

Následující text, vypovídá o dalších vlastnostech rovnoměrného rozdělení modulo 1, a tím nám usnadní hledání benfordovských matematických struktur.

Tvrzení 65. *Buď dána posloupnost reálných čísel (x_n), $n\in\mathbb{N}$.*

(1) *Jestliže $\lim_{n\to\infty}(x_{n+1}-x_n)=\theta$, kde θ je iracionální číslo, potom (x_n) je rovnoměrně rozdělená modulo 1.*

(2) *Jestliže (x_n) je periodická posloupnost, tedy $x_{n+p}=x_n$ pro některá $p\in\mathbb{N}$ a všechna $n\in\mathbb{N}$, potom $(n\theta+x_n)$ je rovnoměrně rozdělená modulo 1.*

(3) *Posloupnost (x_n) je rovnoměrně rozdělená právě tehdy, když $(x_n+\alpha\log_{10}n)$ je rovnoměrně rozdělená modulo 1 pro všechna $\alpha\in\mathbb{R}$*

(4) *Pokud posloupnost (x_n) je rovnoměrně rozdělená modulo 1 a je neklesající, potom $(\frac{x_n}{\log_{10}n})$ je neomezená.*

Poznámka. Vlastnosti z tvrzení (1) a (2) jsou podrobněji popsány v [KN, Věta 3.3 a 2.1]. Dále (3) z [Ber, Lemma 2.8] a (4) z [BBH, Lemma 2.4.(i)]

Pro další tvrzení mějme z Koksmovy metrické věty, dle [KN, Věta 4.3], dáno: Vlastnosti reálných čísel platí pro *skoro každé* $x\in\langle a,b)$, jestliže existuje množina $B\in\mathcal{B}\langle a,b)$, s $\lambda_{a,b}(N)=0$ taková, že tyto vlastnosti platí pro každé $x\notin B$. Tedy řekneme-li, že vlastnost reálného čísla platí pro *skoro každé* x, potom to znamená, že pravděpodobnost dané vlastnosti platí *skoro jistě*, tedy s pravděpodobností 1 pro každou náhodnou veličinu, která má hustotu pravděpodobnosti.

Tvrzení 66. *Buď f_n spojitě diferencovatelná na $\langle a,b\rangle$ pro všechna $n\in\mathbb{N}$. Jestliže $f'_m-f'_n$ je monotónní a $|f'_m(x)-f'_n(x)|\ge\alpha>0$ pro všechna $m\neq n$, kde α nezávisí na x, m a n, potom $(f_n(x))$ je rovnoměrně rozdělená modulo 1 pro skoro každé $x\in\langle a,b\rangle$.*

Věta 67. *Jestliže a, b, α, β jsou reálná čísla, kde $a\neq 0$ a $|\alpha|>|\beta|$, potom $(\alpha^n a+\beta^n b)$ je benfordovská posloupnost právě tehdy, když $\log_{10}|\alpha|$ je iracionální číslo..*

Důkaz. Za daných podmínek, je $\lim_{n\to\infty} \frac{\beta^n b}{\alpha^n a} = 0$, a proto

$$\lim_{n\to\infty} \big(\log_{10} |1 + \frac{\beta^n b}{\alpha^n a}|\big) = 0 = \lim_{n\to\infty} \big(\log_{10} |\alpha^n a + \beta^n b| - \log_{10} |\alpha^n a|\big).$$

Dle Lemmatu 61 (1) je $(\log_{10} |\alpha^n a + \beta^n b|)$ je rovnoměrně rozdělené modulo 1 právě tehdy, když $(\log_{10} |\alpha^n a|) = (\log_{10} |a| + n \log_{10} |\alpha|)$ je také a to dle Tvrzení 65 (1) platí pro iracionální $\log_{10} |\alpha|$. Pro racionální $\log_{10} |\alpha|$ je $n \log_{10} |\alpha|$ racionální. Potom $\mathcal{F}\big(\log_{10} |a| + n \log_{10} |\alpha|\big)$ nabývá pouze spočetně mnoha hodnot a proto by pousloupnost $(\alpha^n a + \beta^n b)$ nebyla rovnoměrně rozdělená modulo 1. Tedy pro iracionální $\log_{10} |\alpha|$ je posloupnost $(\log_{10} |\alpha^n a + \beta^n b|)$ rovnoměrně rozdělená modulo 1 a užitím Lemmatu 60 posloupnost $(\alpha^n a + \beta^n b)$ je benfordovská. □

Příklad 68. Z Příkladu 47 mějme posloupnost (2^n). Užitím Lemmatu 60 máme $n \log_{10} 2$, a jelikož je $\log_{10} 2$ iracionální, pak užitím Tvrzení 65 (1) nebo Věty 67 je posloupnost (2^n) benfordovská. Stejně tak posloupnosti (0.2^n), (0.5^n), (3^n), (5^n) jsou benfordovské a mnoho dalších.

Příklad 69. Posloupnost (10^n) není benfordovská, protože $\log_{10} 10 = 1$, tedy je racionální. Také si lze všimnout dominující první signifikantní číslice 1. I posloupnosti (0.1^n), $(\sqrt{10}^n)$ nejsou benfordovské.

Příklad 70. Posloupnost $(2^n + 10^n)$ není benfordovská dle Věty 67, jelikož v tomto případě je $\alpha = 10$. Lze si také všimnout dominující první signifikantní číslice 1 v $(2^n + 10^n)$. Posloupnosti $(20^n + 10^n)$ a $(2^n + 1^n)$ jsou benfordovské.

Příklad 71. Mějme posloupnost (nx). Potom dle Tvrzení 66 $f'_n = n$, čili $f'_m - f'_n$ je monotónní a $|f'_m(x) - f'_n(x)| \geq 1 > 0$ kde 1 nezávisí na x, m a n pro $m \neq n$. Pak (nx) je rovnoměrně rozdělená modulo 1 pro *skoro každé* $x \in \langle a, b\rangle$. Mějme tedy posloupnost (10^{nx}), potom dle Věty 60 je tato posloupnost pro *skoro všechna* x benfordovská. Je zřejmé, že pro $x = 1$ nebo $x = 0$ tyto posloupnosti benfordovské nejsou.

Příklad 72. Z Příkladu 62 $(\log_{10} n)$ není rovnoměrně rozdělená modulo 1. Mějme tedy $(\log_{10} |\alpha n|)$, $\forall \alpha \in \mathbb{R}$. Posloupnost $(\log_{10} |\alpha n|)$ je neklesající a $(\frac{\log_{10} |\alpha| n}{\log_{10} n})$ je omezená, jelikož $\frac{\log_{10} |\alpha|}{\log_{10} n} + 1 = O(\log_{10} \alpha)$. Užitím Tvrzení 65 (4) $(\log_{10} |\alpha n|)$ není rovnoměrně rozdělená modulo 1, a proto (αn) není benfordovská posloupnost pro všechna $\alpha \in \mathbb{R}$.

Následující větu si vzhledem k rovnoměrnému rozdělení modulo 1 zavedeme již zde. Nicméně, její využití se více projeví až v následujících částech práce.

Věta 73. *Nechť jsou dány náhodné veličiny X, Y.*

(1) *Jestliže X je rovnoměrně rozdělená modulo 1 a Y je nezávislé na X, potom $X + Y$ je rovnoměrně rozdělená modulo 1.*

(2) *Jestliže $\mathcal{F}(X)$ a $\mathcal{F}(X + \alpha)$ mají stejné rozdělení pro iracionální číslo α, potom X je rovnoměrně rozdělená modulo 1.*

(3) *Jestliže (X_n), pro $n \in \mathbb{N}$, je posloupnost náhodných veličin nezávisle identicky rozdělených a X_1 není čistě diskrétní (tzn. $\mathbb{P}(X_1 \in C) < 1$ pro každou spočetnou množinu $C \subset \mathbb{R}$), potom pro všechna $s \in \langle 0, 1)$*

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathcal{F}\big(\sum_{j=1}^{n} X_j\big) \leq s\Big) = s. \tag{4.16}$$

Tedy $\mathcal{F}\left(\sum_{j=1}^{n} X_j\right)$ se blíží k rovnoměrnému rozdělení modulo 1 pro $n \to \infty$.

Důkaz. Odkazujeme na [BH, Věta 4.13]. □

4.2. **Charakterizace invariancí vzhledem ke změně měřítka.**
Invariance vzhledem ke změně měřítka je jedna z nejznámějších vlastností Benfordova zákona. Řekneme-li, že rozdělení je invariantní vzhledem ke změně měřítka, pak číselné soubory vyhovující stejnému rozdělení by si toto rozdělení měly zachovat i například po změně jednotek z centimetrů na palce nebo z kilogramů na libry.

Tvrzení, že benfordovská pravděpodobnostní míra je invariantní vzhledem ke změně měřítka, by však nebylo správné, jelikož neexistuje takto invariantní pravděpodobnostní míra na $(\mathbb{R}^+, \mathcal{B}^+)$. Existuje pouze jediná náhodná veličina X z reálných čísel, pro kterou je pravděpodobnostní rozdělení takto invariantní, nebo také říkáme, že taková náhodná veličina je invariantní vzhledem ke změně měřítka. Tedy mějme náhodné veličiny X a αX mající takto invariantní rozdělení pro všechna $\alpha > 0$, potom X je náhodná veličina taková, že s pravděpodobností 1 je rovna 0, tzn. $\mathbb{P}(X = 0) = 1$. Zřejmě potom $\alpha X = X$. K tomu, aby jsme potvrdili, že je to jediná náhodná veličina s invariantním rozdělením vzhledem ke změně měřítka, mějme náhodnou veličinu X takovou, že $\mathbb{P}(|X| > c) = \delta > 0$, pro $c > 0$. Potom pro $\alpha > 0$ $\lim_{\alpha \to 0^+} \mathbb{P}(|\alpha X| > c) = \mathbb{P}(|X| = \frac{c}{\alpha}) = 0$ a tedy pro dostatečně malá α, kdy $\mathbb{P}(|\alpha X| > c) < \delta = \mathbb{P}(|X| > c)$ to vede ke sporu s invariancí vzhledem ke změně měřítka. Žádná nenulová náhodná veličina tudíž nemůže být invariantní vzhledem ke změně měřítka. Nicméně kladné náhodné veličiny mohou mít rozdělení signifikantní číslice invariantní ke změně měřitka a jak se ukáže, pak nutně musí být benfordovské. Invarianci rozdělení signifikantních číslic ke změně měřítka budeme zkráceně zapisovat *i.r.s.m.*

Definice 74. Pravděpodobnostní míra P na $(\mathbb{R}^+, \mathcal{A})$, s $\mathcal{A} \supset \mathcal{S}$ je *i.r.s.m.*, jestliže pro všechna $\alpha > 0$ a všechny jevy $A \in \mathcal{S}$,

$$P(\alpha A) = P(A),$$

nebo pro všechna $m \in \mathbb{N}$, $\forall d_1 \in \{1, 2, \ldots, 9\}$, $\forall d_j \in \{0, 1, 2, \ldots, 9\}$, $2 \le j \in \mathbb{N}$,

$$P(\{x : D_i(\alpha x) = d_i, i = 1, 2, \ldots, m\}) = P(\{x : D_i(x) = d_i, i = 1, 2, \ldots, m\}).$$

Poznámka. Pozor! Rovnost $P(\alpha A) = P(A)$ neznamená, že $\alpha A = A$.

Příklad 75. Mějme benfordovské pravděpodobnostní rozdělení $\mathbb{B}$ na $(\mathbb{R}^+, \mathcal{S})$. Potom pro všechna $A \in \mathcal{S}$ platí: $A = \bigcup_{k \in \mathbb{Z}} 10^k \langle a, b \rangle$ pro všechna $1 \le a < b < 10$ a pro všechna $\alpha > 0$, $\alpha A = \bigcup_{k \in \mathbb{Z}} \alpha 10^k \langle a, b \rangle = \bigcup_{k \in \mathbb{Z}} 10^{k + \log_{10} \alpha} \langle a, b \rangle$. Jelikož existuje $k \in \mathbb{Z}$ takové, že $\log_{10} \alpha = k + \mathcal{F}(\log_{10} \alpha)$ můžeme psát

$$\alpha A = \bigcup_{k \in \mathbb{Z}} 10^{k + \mathcal{F}(\log_{10} \alpha)} \langle a, b \rangle = \bigcup_{k \in \mathbb{Z}} 10^k C,$$

kde C je dáno jako

$$C = \begin{cases} \langle 10^{\mathcal{F}(\log_{10} \alpha)} a, 10^{\mathcal{F}(\log_{10} \alpha)} b \rangle & \text{pro } 10^{\mathcal{F}(\log_{10} \alpha)} a < 10^{\mathcal{F}(\log_{10} \alpha)} b < 10 \\ \langle 1, 10^{\mathcal{F}(\log_{10} \alpha) - 1} b \rangle \cup \langle 10^{\mathcal{F}(\log_{10} \alpha)} a, 10) & \text{pro } 10^{\mathcal{F}(\log_{10} \alpha)} a < 10 \le 10^{\mathcal{F}(\log_{10} \alpha)} b \\ \langle 10^{\mathcal{F}(\log_{10} \alpha) - 1} a, 10^{\mathcal{F}(\log_{10} \alpha) - 1} b \rangle & \text{pro } 10 \le 10^{\mathcal{F}(\log_{10} \alpha)} a < 10^{\mathcal{F}(\log_{10} \alpha)} b \end{cases}.$$

Potom z Definice 57 pro C ve stejném pořadí

$$\mathbb{B}(\alpha A) = \mathbb{B}(\bigcup_{k\in\mathbb{Z}} 10^k C) = \begin{cases} \log_{10}(10^{\mathcal{F}(\log_{10}\alpha)}b) - \log_{10}(10^{\mathcal{F}(\log_{10}\alpha)}a) \\ \log_{10}(10^{\mathcal{F}(\log_{10}\alpha)-1}b) + 1 - \log_{10}(10^{\mathcal{F}(\log_{10}\alpha)}a) \\ \log_{10}(10^{\mathcal{F}(\log_{10}\alpha)-1}b) - \log_{10}(10^{\mathcal{F}(\log_{10}\alpha)-1}a) \end{cases}$$

$$= \log_{10} b - \log_{10} a = \mathbb{B}(A),$$

tudíž $\mathbb{B}$ je *i.r.s.m.*

Příklad 76. Mějme Dirakovu pravděpodobnostní míru δ_1 a $A \in \mathscr{S}$ takové, že $A = \bigcup_{k\in\mathbb{Z}} 10^k\langle 1,2)$. Potom můžeme psát

$$\begin{aligned} \delta_1(A) = \delta_1(\bigcup_{k\in\mathbb{Z}} 10^k\langle 1,2)) &= \delta_1(D_1 = 1) = 1 \neq \\ &\neq 0 = \delta_1(D_1 \in \{2,3\}) = \delta_1(\bigcup_{k\in\mathbb{Z}} 10^k\langle 2,4)) = \delta_1(2A). \end{aligned}$$

Jelikož $\exists\alpha > 0$ takové, že $\delta_1(A) \neq \delta_1(\alpha A)$, potom δ_1 není *i.r.s.m.*

Příklad 77. Mějme náhodnou veličinu X na $(\langle 0,1), \mathcal{B}\langle 0,1), \lambda_{0,1})$. Potom

$$\lambda_{0,1}(D_1(X) = 1) = \frac{1}{9} < \frac{10}{18} = \lambda_{0,2}(D_1(2X) = 1)$$

a proto náhodná veličina X není *i.r.s.m.*.

Poznámka. Náhodná veličina $2X$ je na $(\langle 0,2), \mathcal{B}\langle 0,2), \lambda_{0,2})$, na intervalu $\langle 0,1)$ je čísel začínající signifikantní číslicí 1 stále $1/9$, na intervalu $\langle 1,2)$ jsou to všechna taková čísla. Intervaly jsou však celkově poloviční ku celkovému intervalu, proto $\lambda_{0,2}(D_1(2X) = 1) = 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{9} = \frac{10}{18}$.

Věta 78. *Pravděpodobnostní míra P na $(\mathbb{R}^+, \mathcal{A})$, s $\mathcal{A} \supset \mathscr{S}$ je i.r.s.m. právě tehdy, když pro všechna $A \in \mathscr{S}$, $P(A) = \mathbb{B}(A)$, tedy právě tehdy, když P je benfordovská pravděpodobnostní míra.*

Důkaz. Mějme pravděpodobnostní míru $P(\mathbb{R}^+, \mathscr{S})$, a z Lemmatu 58 mějme $Q = \ell_* P$ a tudíž Q je pravděpodobnostní míra na $(\langle 0,1), \mathcal{B}\langle 0,1))$. Jestliže P je *i.r.s.m.*, potom pro všechna $\alpha > 0$ a všechna $A \in \mathcal{A}$, $P(\alpha A) = P(A)$ je ekvivalentní s

$$Q(\mathcal{F}(t + B)) = Q(B) \tag{4.17}$$

pro všechna $t \in \mathbb{R}$ a každé $B \in \mathcal{B}\langle 0,1)$, kde $\mathcal{F}(t+B) = \{\mathcal{F}(t+B) : \ x \in B\}$. Mějme tedy náhodnou veličinu X s pravděpodobnostním rozdělením Q. Potom $Q(\mathcal{F}(t+X)) = Q(X)$ podle Věty 73 (1) a (2) může nastat právě tehdy, když X je rovnoměrně rozdělená modulo 1 a tedy Q na $(\langle 0,1), \mathcal{B}\langle 0,1))$ je rovnoměrné rozdělení $\lambda_{0,1}$. (Připomeňme, že konstantní náhodná veličina, v našem případě t, je nezávislá vůči všem ostatním náhodným veličinám). Tedy $P(A) = \lambda_{0,1}(\ell(A)) = \lambda_{0,1}(\log_{10} S(A))$ a z Věty 60 $\lambda_{0,1}(\log_{10} S(A)) = \lambda_{0,1}(\mathcal{F}(\log_{10} A)) = \mathbb{B}(A)$. □

Nyní obdobně popíšeme kladné reálné funkce a posloupnosti reálných čísel.

Definice 79. Posloupnost reálných čísel (x_n) je *i.r.s.m.*, jestliže pro všechna $\alpha > 0$ a všechna $t \in \langle 1,10)$ platí rovnost

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : \ S(\alpha x_n) < t\}}{N} = \lim_{N\to\infty} \frac{\#\{1 \le n \le N : \ S(x_n) < t\}}{N}. \tag{4.18}$$

Borelovsky měřitelná funkce $f : \langle 0, \infty) \to \mathbb{R}$ je *i.r.s.m.*, jestliže pro všechna $\alpha > 0$ a všechna $t \in \langle 1, 10)$ platí

$$\lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0, T) :\ S(\alpha f(\tau)) < t\})}{T} = \lim_{T\to\infty} \frac{\lambda(\{\tau \in \langle 0, T) :\ S(f(\tau)) < t\})}{T}$$

..

K následující větě si řekněme, že množina $A \subset \mathbb{N}$ má *hustotu* $\rho \in \langle 0, 1\rangle$, jestliže existuje limita taková, že

$$\lim_{N\to\infty} \frac{\#\{1 \leq n \leq N :\ n \in A\}}{N} = \rho.$$

Například $\rho(\{n :\ n \text{ je liché}\}) = \frac{1}{2}$, naopak pro $\{n :\ D_1(n) = 2\}$ limita neexistuje.

Věta 80.

(1) *Buď dána posloupnost reálných čísel* (x_n) *a nechť* $\{n :\ x_n \neq 0\} = \{n_1 < n_2 < n_3 < \dots\}$. *Potom posloupnost* (x_n) *je i.r.s.m. právě tehdy, když* $\{n :\ x_n \neq 0\}$ *má hustotu a buďto* $\rho(\{n :\ x_n = 0\}) = 1$ *a nebo* $(x_{n_j}) = \{x_n :\ x_n \neq 0\}$, $j \in \mathbb{N}$, *je benfordovská posloupnost. V případě, kdy* $\rho(\{n :\ x_n = 0\}) = 0$, *posloupnost* (x_n) *je i.r.s.m. právě tehdy, když je posloupnost benfordovská.*

(2) *Borelovsky měřitelná funkce* $f : \langle 0, \infty) \to \mathbb{R}$ *s* $\lambda(\{t > 0 :\ f(t) = 0\}) < \infty$ *je i.r.s.m., když je funkce benfordovská.*

Důkaz. Odkazujeme na [BH, Věta 4.22]. □

Příklad 81. Mějme posloupnost (2^n) z Příkladu 47 o které víme, že je benfordovská a nemá nulové prvky, potom dle Věty 80 je *i.r.s.m.* To, že pro $\alpha > 0$ je $(\alpha 2^n)$ také benfordovská posloupnost, je zřejmé z Věty 67.

Ve Větě 78 jsme si ukázali, že číselný soubor, který vyhovuje Benfordovu zákonu, je nutně *i.r.s.m.* Nicméně Steven W. Smith ve své knize [Sm] pozoruje, že postačující podmínkou k tomu, aby pravděpodobnostní míra P na $(\mathbb{R}^+, \mathcal{B}^+)$, která je *i.r.s.m.*, byla benfordovská, je aby takto invariantní byla jen vůči jediné signifikantní číslici. Ačkoliv toto platí pro jakoukoliv signifikantní číslici, pro jednoduchost v následující větě byla zvolena pouze první signifikantní číslice.

Věta 82. *Pro každou náhodnou veličinu* X *takovou, že* $\mathbb{P}(X = 0) = 0$, *jsou následující tvrzení ekvivalentní.*

(1) X *je benfordovská náhodná veličina*

(2) *Existuje číslo* $d_1 \in \{1, 2, \dots, 9\}$ *takové, že pro všechna* $\alpha > 0$

$$\mathbb{P}(D_1(\alpha X) = d_1) = \mathbb{P}(D_1(X) = d_1),$$

kde $\mathbb{P}(D_1(X) = d_1) = \log_{10}(1 + d_1^{-1})$.

Důkaz. Odkazujeme na [BH, Věta 4.25]. □

Testování, jestli číselný soubor vyhovuje Benfordovu zákonu skrze *i.r.s.m.*, by nemuselo být příliš lákavé. Při našich znalostech do teď získaných informacích k *i.r.s.m.*, bychom museli číselný soubor přenásobit všemi $\alpha > 0$ a pak porovnávat relativní četnosti všech čísel, jejichž první signifikantní číslice je rovná například 1, abychom si byli jisti, že vyhovuje Benfordovu zákonu. Nicméně, díky zjištění S. W. Smitha v [Sm] je postačující, abychom číselný soubor testovali přenásobením posloupností (α^n), kde $\log_{10}\alpha$ je iracionální. Tedy jestliže je náhodná veličina X benfordovská, potom pro posloupnost (α^n) pro pevně dané $d_1 \in \{1, 2, \ldots, 9\}$ a všechna $n \in \mathbb{N}$ platí:

$$\mathbb{P}(D_1(\alpha^n X) = d_1) = \mathbb{P}(D_1(X) = d_1).$$

Část 5. Diferenciální rovnice

V této části práce si krátce představíme Benfordův zákon v souvislosti s časově spojitě deterministickými procesy. Deterministické procesy jsou takové procesy, které jsou určeny počátečním stavem, a každý následující krok je závislý na svém předchozím. I když se v tomto smyslu dá například hovořit i o iteracích zobrazení a mocninách matic, v našem případě jsme si zvolili hovořit o diferenciálních rovnicích. Pro více informací ohledně Benfordova zákona v souvislosti s deterministickými procesy odkazujeme na [BH, str. 64].

Mějme tedy Cauchyho úlohu

$$y' = F(y), \quad y(0) = y_0, \tag{4.19}$$

kde $y_0 \in \mathbb{R}$, $F : \mathbb{R} \to \mathbb{R}$ je spojitě diferencovatelná a $F(0) = 0$. Nejjednodušší případ nastane, když $F(y) = \alpha y$ pro některé $\alpha \in \mathbb{R}$, a tedy řešením námi dané Cauchyho úlohy je $y(t) = y_0 e^{\alpha t}$. Jestliže $\alpha y_0 \neq 0$, pak toto řešení je dle Příkladu 63 benfordovské.

Lemma 83. *Nechť jsou dány Borelovsky měřitelné funkce $f, g : \langle 0, \infty) \to \mathbb{R}$ takové, že $|f(t)| \to \infty$, $|g(t)| \to \infty$ pro $t \to \infty$ a $\sup_{t \geq 0} |f(t) - g(t)| < \infty$. Potom f je benfordovská funkce právě tehdy, když g je také benfordovská funkce.*

Důkaz. Mějme konstantu c danou jako $c = \sup_{t \geq 0} |f(t) - g(t)| + 1$. Na intervalu typu $\langle \alpha, \infty)$ mějme $|f(t)|, |g(t)| > c$. Potom platí

$$\begin{aligned} -\log_{10}\left(1 + \frac{c}{|f(t)| - c}\right) &\leq \log_{10} \frac{|g(t)|}{|g(t)| + c} \\ &\leq \log_{10} \frac{|g(t)|}{|f(t)|} \\ &\leq \log_{10} \frac{|f(t)| + c}{|f(t)|} \leq \log_{10}\left(1 + \frac{c}{|f(t)| - c}\right). \end{aligned}$$

Jelikož

$$\left|\log_{10} |g(t)| - \log_{10} |f(t)|\right| = \left|\log_{10} \frac{|g(t)|}{|f(t)|}\right| \leq \log_{10}\left(1 + \frac{c}{|f(t)| - c}\right)$$

a

$$\lim_{t \to \infty} \log_{10}\left(1 + \frac{c}{|f(t)| - c}\right) = 0,$$

potom dle Lemmatu 61 (1) je $\log_{10} |f(t)|$ rovnoměrně rozdělené modulo 1 právě tehdy, když je $\log_{10} |g(t)|$, a tedy dle Věty 60 je funkce $f(t)$ benfordovská právě tehdy, když je $g(t)$ také. □

Věta 84. *Buď dána funkce $F : \mathbb{R} \to \mathbb{R}$, která je C^2, $F(0) = 0$ a $F'(0) < 0$. Potom pro každé $y_0 \neq 0$, které je zároveň dostatečně blízko 0, je řešení Cauchyho úlohy (4.19) benfordovské.*

Důkaz. Mějme δ větší než nula a zároveň dostatečně blízké 0, aby platilo $yF(y) < 0$ pro všechna $0 < |y| \leq \delta$. Potom pro námi dané F má Cauchyho úloha (4.19) lokální řešení kdykoliv $|y_0| \leq \delta$. Říkáme, že $\langle -\delta, \delta \rangle$ je tzv. dopředně invarientní. Nechť tedy $y_0 \in (0, \delta\rangle$ a mějme řešení Cauchyho úlohy (4.19) dáno jako $y = y(t)$. Potom $\lim_{t \to \infty} y(t) = 0$. Mějme dáno $z : \langle 0, \infty) \to \mathbb{R}$ definované jako $z = y^{-1}$, tedy $z(0) = z_0 = y_0^{-1}$ a $\lim_{t \to \infty} |z(t)| = \infty$. Buď $\alpha = -F'(0) > 0$ a mějme korektně definovanou spojitou funkci $g : \mathbb{R} \to \mathbb{R}$, pro kterou $F(y) = -\alpha y + y^2 g(y)$. Potom z

$$z' = -\frac{y'}{y^2} = \alpha z - g(z^{-1}),$$

užitím metody variace konstant, pro $t \geq 0$,

$$z(t) = \mathrm{e}^{\alpha t} z_0 - \int_0^t \mathrm{e}^{\alpha(t-\tau)} g\big(z(\tau)^{-1}\big)\ \mathrm{d}\tau.$$

Pro $\alpha > 0$ a spojitou funkci g mějme pomocné číslo

$$\overline{z_0} = z_0 - \int_0^\infty \mathrm{e}^{-\alpha\tau} g\big(z(\tau)^{-1}\big)\ \mathrm{d}\tau.$$

Potom pro $t \geq 0$ a užitím substituce

$$\begin{aligned} |z(t) - \mathrm{e}^{\alpha t}\overline{z_0}| &= \Big|\int_t^\infty \mathrm{e}^{\alpha(t-\tau)} g\big(z(\tau)^{-1}\big)\ \mathrm{d}\tau\Big| \\ &\leq \int_0^\infty \mathrm{e}^{-\alpha\tau} |g\big(z(t+\widetilde{\tau})^{-1}\big)|\ \mathrm{d}\widetilde{\tau} \leq \frac{\max_{|y|\leq\delta} |g(y)|}{\alpha}. \end{aligned}$$

Tedy $\sup_{t\geq 0} |z(t) - \mathrm{e}^{\alpha t}\overline{z_0}| < \infty$ a užitím Lemmatu 83 je z benfordovská funkce, když je $\mathrm{e}^{\alpha t}\overline{z_0}$ také. Dle Příkladu 63 $\mathrm{e}^{\alpha t}\overline{z_0}$ benfordovské je a jelikož $y = z^{-1}$, pak užitím (2) z Důsledku (64) je y benfordovská funkce. □

Příklad 85. Funkce $F(y) = -y + y^4\mathrm{e}^{-y^2}$ splňuje podmínky z Věty 84. Tedy F je C^2, $F(0) = 0$ a jelikož $F'(y) = -1 - \mathrm{e}^{-2y}y^4 + 4y^3\mathrm{e}^{-y^2}$, pak $F'(0) < 0$. Tudíž řešení Cauchyho úlohy (4.19) je pro všechna $y \neq 0$ benfordovské.

Příklad 86. Mějme funkci $F(y) = -y^3 + y^4\mathrm{e}^{-y^2}$, potom nesplňuje všechny podmínky z Věty 84 jelikož $F'(0) = 0$. Abychom si tedy potvrdili, že řešení Cauchyho úlohy (4.19) není benfordovské, mějme $y(0) = y_0 \neq 0$ a k získání určitého implicitního tvaru integrujme

$$\begin{aligned} \int_0^t -\frac{y'}{y^3}\ \mathrm{d}t &= \int_0^t 1 - y\mathrm{e}^{-y^2}\ \mathrm{d}t, \\ \frac{1}{2y^2} + c &= t - \int_0^t y\mathrm{e}^{-y^2}\ \mathrm{d}t, \\ \frac{1}{2y^2} - \frac{1}{2y_0^2} &= t - \int_0^t y\mathrm{e}^{-y^2}\ \mathrm{d}t, \\ \frac{\frac{y_0^2}{y^2} - 1}{2y_0^2} &= t - \int_0^t y\mathrm{e}^{-y^2}\ \mathrm{d}t, \end{aligned}$$

$$y(t)^2 = \frac{y_0^2}{1 + 2ty_0^2 - 2y_0^2\int_0^t y(\tau)\mathrm{e}^{-y(\tau)^2}\ \mathrm{d}\tau}. \tag{4.20}$$

O funkci F víme, že je hladká, a $yF(y) < 0$ pro všechna $y \neq 0$. Tedy pro všechny $y_0 \in \mathbb{R}$ má Cauchyho úloha (4.19) řešení, jež $\lim_{t\to\infty} y(t) = 0$. Tudíž také $\lim_{t\to\infty} \frac{1}{t}\int_0^t y(\tau)\mathrm{e}^{-y(\tau)^2}\ \mathrm{d}\tau = 0$ a potom z (4.20) $\lim_{t\to\infty} 2ty(t)^2 = 1$. Pak $|\log_{10} y|$ je neklesající a $t \to \frac{|\log_{10} y|}{\log_{10} t}$ je omezené, a tedy pro spojitě upravené (4) v Tvrzení 65 není $|\log_{10} y|$ rovnoměrně rozdělené modulo jedna a tudíž není funkce y benfordovská.

Příklad 87. Mějme funkci $F(y) = -y + y\log_{10}(1 + y^2)$. Funkce F má 3 kořeny, $y = 0$ a $y = \pm 3$. Dle Věty 84 je řešení Cauchyho úlohu (4.19) benfordovské pro $0 < |y_0| < 3$. Pro intervaly $(-\infty, 3)$ a $(3, \infty)$ mějme $|y_0| > 3$. Potom mějme $z = \log_{10} y - \frac{1}{2}$ a tedy

$$z' = \frac{y'}{y \ln 10} = \frac{2z}{\ln 10} + \frac{\log_{10}(1 + 10^{-1-2z})}{\ln 10}.$$

Pak pro všechna $t \geq 0$

$$z(t) = e^{2t/\ln 10} z_0 + \int_0^t e^{2(t-\tau)/\ln 10} \frac{\log_{10}(1 + 10^{-1-2z(\tau)})}{\ln 10}\, d\tau$$

a pomocné číslo

$$\overline{z_0} = z_0 + \int_0^\infty e^{-2\tau/\ln 10} \frac{\log_{10}(1 + 10^{-1-2z(\tau)})}{\ln 10}\, d\tau.$$

Potom

$$\begin{aligned} |z(t) - e^{2t/\ln 10}\overline{z_0}| &= \Big| \int_t^\infty e^{2(t-\tau)/\ln 10} \frac{\log_{10}(1 + 10^{-1-2z(\tau)})}{\ln 10}\, d\tau \Big| \\ &\leq \int_0^\infty e^{-2\tau/\ln 10} \frac{\log_{10}(1 + 10^{-1-2z(t+\widetilde{\tau})})}{\ln 10}\, d\widetilde{\tau}. \\ &\leq \log_{10} \sqrt{1 + 10^{-1-2z(t)}}. \end{aligned}$$

Jelikož $\lim_{t\to\infty} \log_{10} \sqrt{1 + 10^{-1-2z(t)}} = 0$, pak $z(t)$ je rovnoměrně rozdělená modulo 1 právě tehdy, když je i $t \to e^{2t/\ln 10}\overline{z_0}$. Jelikož $t \to e^{2t/\ln 10}\overline{z_0}$ podle [BH, Příklad 4.5 (iii)] je rovnoměrně rozdělená modulo 1, pak tedy dle $z = \log_{10} y - \frac{1}{2}$ je y benfordovská funkce pro $|y_0| > 3$.

Část 6. Konvergence k Benfordovu zákonu

V této části se skrze konvergenci k Benfordovu zákonu pokusíme nastínit, proč se soubory přirozeně se vyskytujících čísel tak často „blíží“ k Benfordovu zákonu.

Připomeňme, že posloupnost náhodných veličin (X_n) konverguje v rozdělení k náhodné veličině X, jestliže $\lim_{n\to\infty} \mathbb{P}(X_n \leq t) = \mathbb{P}(X \leq t)$ platí pro všechny $t \in \mathbb{R}$, pro které $\mathbb{P}(X = t) = 0$. Tuto konvergenci značíme $X_n \xrightarrow{\mathcal{D}} X$. Menší úpravou řekněme, že (X_n) konverguje v rozdělení k benfordovské náhodné veličině X, jestliže $S(X_n) \xrightarrow{\mathcal{D}} S(X)$. Tedy, že pro všechna $t \in \langle 1, 10)$ platí, $\lim_{n\to\infty} \mathbb{P}(S(X_n) \leq t) = \log_{10} t$. Také řekněme, že posloupnost náhodných veličin (X_n), konverguje k náhodné veličině X *skoro jistě* (*s.j.*), jestliže $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.

Nejlepším způsobem jak představit posloupnosti náhodných veličin ve smyslu Benfordova zákona, je zvolit $X_n = X^n$pro každé $n \in \mathbb{N}$. Je zřejmé, že posloupnost (X^n) není nezávisle identicky rozdělená v případě, kdy $X \notin \{0, 1\}$.

Věta 88. *Mějme náhodnou veličinu X s hustotou f_X. Potom*

(1) *(X^n) konverguje v rozdělení k benfordovské náhodné veličině.*
(2) *S pravděpodobností 1 je (X^n) benfordovská posloupnost náhodných veličin.*

Důkaz. Odkazujeme na [BH, Věta 6.1]. □

Pro jakoukoliv náhodnou veličinu X a tedy tvrzení (2) z Věty 88 platí, jestliže platí tvrzení (1). Pro nezávisle identicky rozdělenou posloupnost náhodných veličin (X_n) konvergence v rozdělení k benfordovské náhodné veličině znamená, že X_n je benfordovská náhodná veličina pro všechny $n \in \mathbb{N}$, a tedy dle nezávislosti náhodných veličin je (X_n) benfordovská posloupnost náhodných veličin s pravděpodobností jedna. Nicméně obecně jsou tyto vlastnosti nezávislé.

Příklad 89. Mějme X rovnoměrně rozdělenou na $\langle 0, 1)$. Potom pro všechna $n \in \mathbb{N}$ a všechna $t \in \langle 1, 10)$ platí
$$F_{S(X^n)}(t) = \frac{t^{1/n} - 1}{10^{1/n} - 1}.$$
Pak s vědomím $\frac{e^t-1-t}{e^t-1} < \frac{t}{2}$ pro všechna $t > 0$ a $n \to \infty$
$$|F_{S(X^n)}(t) - \log_{10} t| \leq \frac{10^{1/n} - 1 - \frac{\ln 10}{n}}{10^{1/n} - 1} < \frac{\ln 10}{2n} \to 0.$$
Tudíž (X^n) konverguje v rozdělení k benfordovské náhodné veličině a jelikož $\log_{10} X$ je iracionální *skoro jistě* pak (X_n) je benfordovská s pravděpodobností 1.

Příklad 90. Předpokládejme, že $X = 2$ *skoro jistě.* Tudíž X nemá hustotu. Pro každé $n \in N$ platí, že $S(X^n) = 10^{\mathcal{F}(n \log_{10} 2)}$ s pravděpodobností 1, tedy (X^n) nekonverguje v rozdělení k benfordovské náhodné veličině. Na druhou stranu, to že je (X^n) benfordovská posloupnost *skoro jistě,* lze snadno odvodit z Příkladu 71. Víme, že $\log_{10} 2$ je iracionální číslo, a $10^{\mathcal{F}(n \log_{10} 2)} = \frac{10^{n \log_{10} 2}}{10^{\lfloor n \log_{10} 2 \rfloor}}$, tedy $(10^{\mathcal{F}(n \log_{10} 2)})$ má stejné signifikantní číslice jako $(10^{n \log_{10} 2})$.

Následující věta na rozdíl od Věty 88 bere v potaz součiny nezávislých náhodných veličin.

Věta 91. *Mějme nezávislé náhodné veličiny X a Y s $\mathbb{P}(XY = 0) = 0$. Potom*

(1) *jestliže X je benfordovská náhodná veličina, potom je i XY.*

(2) *Jestliže $S(X)$ a $S(XY)$ mají stejné rozdělení, potom buď $\log_{10} S(Y)$ je racionální s pravděpodobností 1, nebo X je benfordovská náhodná veličina.*

Důkaz. Odkazujeme na [BH, Věta 6.3]. □

Příklad 92. Mějme nezávislé náhodné veličiny W a Z s rovnoměrným rozdělením na $(0, 1)$. Potom také $X = 10^W$, $Y = Z$ jsou nezávislé a dle (1) z Věty 91 je XY benfordovská náhodná veličina. Jestliže je dáno $X = 10^W$ a $Y = 10^{1-W}$, pak jsou obě náhodné veličiny benfordovské. Náhodná veličina $XY = 10$ ale není benfordovská, a tudíž nezávislost je nutnou součástí Věty 91 (1).

Příklad 93. Mějme náhodnou veličinu X rovnou buď $10^{\sqrt{2}-1}$ nebo $10^{2-\sqrt{2}}$ s pravděpodobností $\frac{1}{2}$ a nechť $Y = X^{-2}$. I když $S(X)$ a $S(XY) = S(X^{-1})$ mají stejné rozdělení, X není benfordovská náhodná veličina a $\log_{10} Y$ není racionální. Tudíž z toho plyne podmínka nezávislosti i pro tvrzení (2) z Věty 91.

Důsledek 94. *Nechť X je náhodná veličina s $\mathbb{P}(X = 0) = 0$, a nechť α je iracionální číslo. Jestliže $S(X)$ a $S(\alpha X)$ mají stejné rozdělení, potom X je benfordovská náhodná veličina.*

Nyní uvedeme nejdůležitější větu, vypovídající o Benfordovu zákonu, již tvrzení je možnou příčinou výskytu rozdělení signifikantních číslic dané Benfordovým zákonem.

Věta 95. *Buď (X_n) nezávisle identicky rozdělená posloupnost náhodných veličin, které nejsou čistě atomické, tzn. $\mathbb{P}(X_1 \in C) < 1$ pro každou spočetnou množinu $C \subset \mathbb{R}$. Potom*

(1) *$\left(\prod_{j=1}^{n} X_j\right)$ konverguje v distribuci k Benfordovu zákonu*
(2) *S pravděpodobností 1 $\left(\prod_{j=1}^{n} X_j\right)$ je posloupnost benfordovská.*

Důkaz. Odkazujeme na [BH, Větu 6.6]. □

Část 7. Identifikace benfordovských souborů

V předchozích částech práce jsme si ukázali několik specifických vlastností Benfordova zákona. Nyní tyto vědomosti budeme aplikovat na Fibonacciho posloupnost a posloupnost prvočísel ze začátku práce. Dále se podíváme na některé soubory „přirozeně“ se vyskytujících čísel, na které aplikujeme Pearsonův χ^2 test dobré shody.

7.1. Aplikace charakterizací Benfordova zákona.

Příklad 96. Podobně jako u příkladu 33 uvažujme Fibonacciho posloupnost (F_n).

(1) *Charakterizace Fibonacciho posloupnosti rovnoměrným rozdělením modulo 1.* Připomeňme vztahy pro n-tý člen Fibonacciho posloupnosti pro všechna $n \in \mathbb{N}$ je dán

$$F_n = \frac{1}{\sqrt{5}}\Big((\frac{1+\sqrt{5}}{2})^n - (\frac{1-\sqrt{5}}{2})^n\Big) = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}},$$

kde $\varphi = \frac{1}{2}(1+\sqrt{5}) = 1.618\ldots$. Ověřme předpoklady Věty 67 s volbami: $\alpha = \varphi$, $\beta^n = -\varphi^{-1}$, $a = \frac{1}{\sqrt{5}} = -b$. Víme, že $\varphi > 1$, tedy platí $|\alpha| > |\beta|$. Jelikož $\log_{10}\varphi$ je iracionální číslo, posloupnost (F_n) je benfordovská.

(2) *Invariance rozdělení signifikantních číslic vzhledem ke změně měřítka.* Budeme postupovat podle věty 80. Fibonacciho posloupnost nemá nulové prvky, a tudíž je benfordovskou posloupnosít právě tehdy, když je *i.r.s.m.* (tzn. pro $\alpha > 0$ má (αF_n) stejné rozdělení signifikantních číslic). Zkoumejme tedy počty prvních signifikantních číslic pro prvních 10^4 prvků Fibonacciho posloupnosti dané následující tabulkou:

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	R_d = [%]
(F_n)	3011	1761	1251	968	792	668	580	513	456	0.016
(πF_n)	3010	1760	1250	970	791	671	579	511	458	0.015
$(F_n/7)$	3012	1761	1249	970	791	670	580	510	457	0.017
$n \cdot \mathbb{B}(D_1 = d_1)$	*3010*	*1761*	*1249*	*969*	*792*	*669*	*580*	*512*	*458*	–

Tabulka 5. Počet prvků přenásobené Fibonacciho posloupnosti, začínajících danou signifikantní číslicí (pro 10^4 prvků).

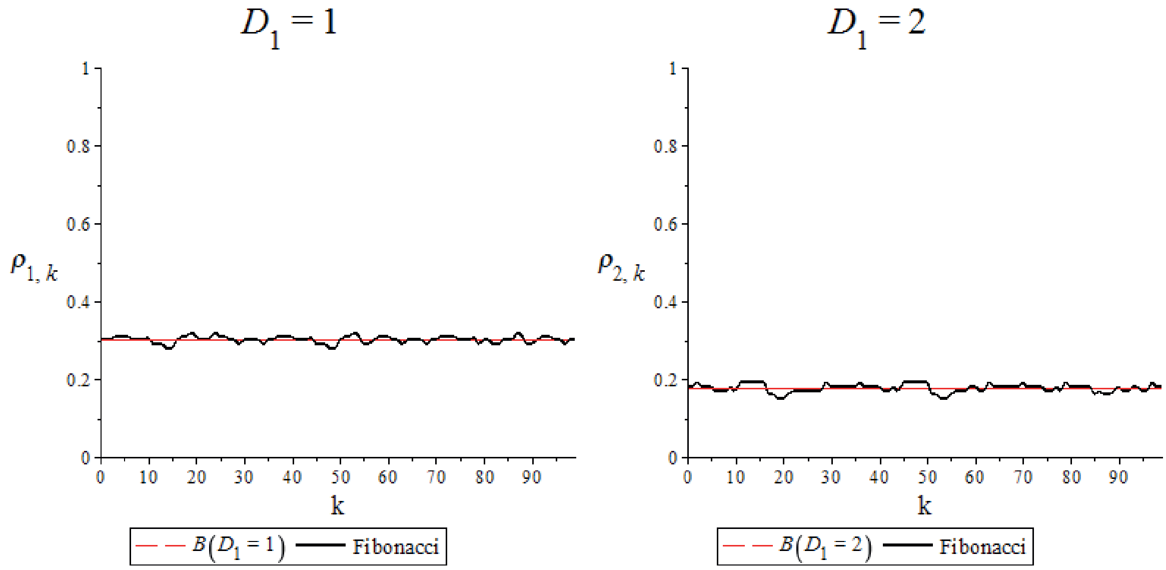
Podobně jako na začátku práce k porovnání shody využili maximální odchylku

$$R_d = \max_{d_1=1\ldots9} |\rho_n(d_1) - \log_{10}(1 + d_1^{-1})|.$$

Nahlédnutím do tabulky si lze si tedy všimnout, že Fibonacciho posloupnost pro 10^4 prvků vykazuje známky *i.r.s.m.*.

(3) *Smithův Ones scale test.* Tento test se zkráceně značí OST a je podrobně popsán v [Sm]. Dle *i.r.s.m.* by číselný soubor, který vyhovuje Benfordovu zákonu, měl zachovat relativní četnost čísel, které začínají danou signifikantní číslicí i po přenásobení prvky zadané geometrické posloupnosti. Na této vlastnosti je založen Smithův test. Aplikujme jej na prvních sto čísel Fibonacciho posloupnost, kterou postupně přenásobujeme prvky posloupnosti (1.07^k), a zdůrazněme, že $\log_{10} 1.07$ je iracionální číslo.

Postupně klademe $k = 1, 2, \ldots, 100$ a počítáme relativní četnosti čísel s první signifikantní číslicí $d_1 = 1$, respektive $d_1 = 2$. Na Obrázku 7.1 jsou znázorněny relativní četnosti $\rho_{d_1,k} = \dfrac{\#\{n \in \mathbb{N} :\ 1 \le n \le 100,\ D_1(1.07^k F_n) = d_1\}}{100}$ odpovídající volbám n a d_1, pro různé volby k. Červeně jsou pak vyznačeny příslušné pravděpodobnosti benfordovské náhodné veličiny. Ke které první signifikantní číslici se graf vztahuje udáváme titulkem grafu.



Obrázek 7.1. Ones scaling test Fibonacciho posloupnosti pro první signifikantní číslici rovnou $d_1 \in \{1, 2\}$.

V porovnání s prvočísly níže se relativní četnosti pro Fibonacciho posloupnost od Benfordova zákona moc neliší.

Příklad 97. Ve stejném smyslu jako Fibonacciho posloupnost mějme posloupnost prvočísel (p_n) z Příkladu 34.

(1) *Charakterizace posloupnosti prvočísel rovnoměrným rozdělením modulo 1.* Z Prvočíselné věty víme $p_n = O(n \log_{10} n)$ pro $n \to \infty$. Posloupnost $(\log_{10} p_n)$ je neklesající, ale $(\frac{\log_{10}(p_n)}{\log_{10} n})$ je omezená, a proto posloupnost $(\log_{10} p_n)$ není rovnoměrně rozdělená modulo 1 dle Věty 65. Tudíž (p_n) není benfordovská.

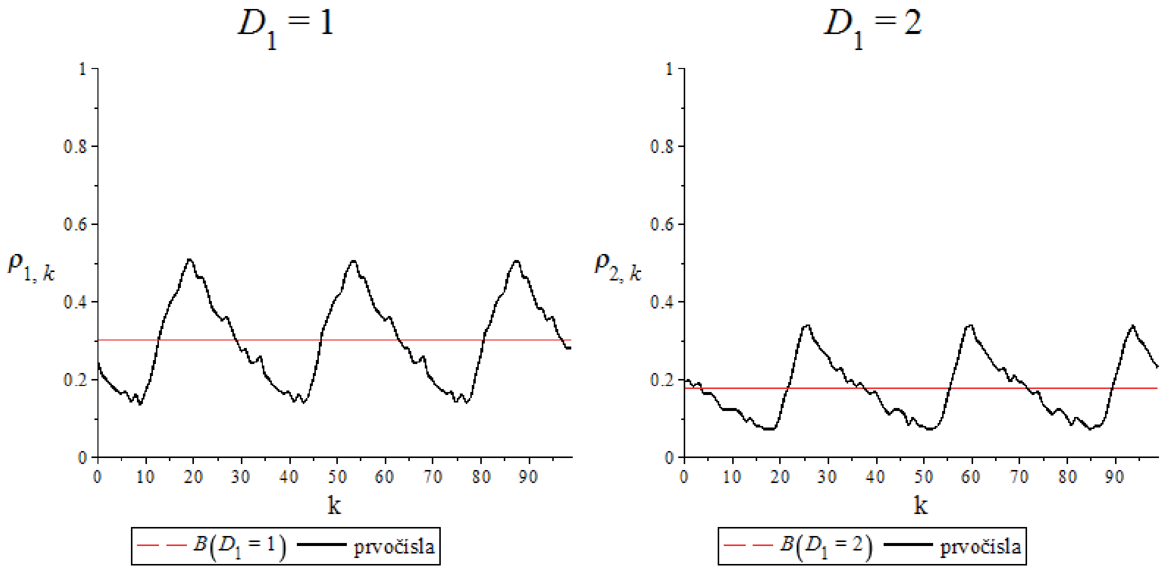
(2) *Charakterizace posloupnosti prvočísel skrze i.r.s.m.* .

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$R_d = [\%]$
(p_n)	1601	1129	1097	1069	1055	1013	1027	1003	1006	14.09
(πp_n)	3398	3231	1182	388	363	372	361	362	343	14.70
$(p_n/7)$	3932	818	799	773	758	737	743	732	708	9.43
$n \cdot \mathbb{B}(D_1 = d_1)$	*3010*	*1761*	*1249*	*969*	*792*	*669*	*580*	*512*	*458*	–

Tabulka 6. Počet prvků přenásobené posloupnosti prvočísel začínajících danou signifikantní číslicí, (pro 10^4 prvků).

Pro naše vybraná $\alpha = \{\pi, \frac{1}{7}\}$ se počet prvních signifikantních číslic v (αp_n) dosti liší jak mezi sebou, tak od původní posloupnosti prvočísel nebo od Benfordova zákona.

(3) *Smithův Ones scale test posloupnosti prvočísel.* Stejně jako u Fibonacciho posloupnosti mějme prvních $n = 100$ prvků prvočísel, kterou přenásobíme posloupností (1.07^k) pro $k = 100$. Prvočíselnou posloupnost opět testujeme na první signifikantní číslici rovnou $d_1 \in \{1, 2\}$



Obrázek 7.2. Ones scaling test posloupnosti prvočísel pro první signifikantní číslici rovnou $d_1 \in \{1, 2\}$.

Tedy lze si všimnout že maximální odchylka všech relativních četností od „teoretické relativní četnosti" dané Benfordovým Zákonem je větší než u Fibonacciho čísel.

7.2. **Empirická pozorování.**

V dnešní době existuje mnoho způsobů jak získat některé číselné soubory. Nejčastěji na stránkách statistických úřadů jednotlivých států, kde najdeme číselné soubory týkajících se populace, úmrtí, rozlohy, volby, ale i jiné. Mnoho takových souborů už bylo testováno na benfordovské rozdělení pravděpodobnostních číslic. V práci proto volíme číselné soubory, které se ve spojení s Benfordovým zákonem tak často neobjevují, a v dnešní době je můžeme považovat za zajímavé. Nebudeme však volit číselné soubory, u kterých je na první pohled zřejmé, že Benfordovu zákonu vyhovovat nebudou, jako například údaje o vzrůstu dospělé populace, které s velkou pravděpodobností budou začínat 1.

K zjištění jak moc vyhovuje námi zvolený číselný soubor Benfordovu zákonu se nejvíce hodí Pearsonův χ^2 test dobré shody. Tento test popisuje, na jaké úrovni se námi pozorované hodnoty, námi nalezený číselný soubor, shodují s očekávanými hodnotami, námi zvoleným teoretickým rozdělením pro n počet prvků. Předpoklad, že pozorované hodnoty se blíží očekávaným hodnotám, je dán nulovou hypotézou H_0, kterou zamítám, nebo nezamítáme. Při zamítnutí nulové hypotézy tvrdíme, že se pozorované hodnoty neřídí dle zvoleného rozdělení. Při nezamítnutí H_0 předpoklad stále platí, to ale neznamená, že číselný soubor se určitě řídí dle očekávaných hodnot.

Test budeme provádět na hladině významnosti $\alpha = 0.05$, tedy existuje 5% šance, že hypotéza je zamítnuta neprávem. Dle Pearsonova χ^2 testu dobré shody určíme p-hodnotu, která je dána jako nejnižší hladina významnosti, na které můžeme hypotézu zamítnout. Pro $\alpha \geq p$-hodnota hypotézu zamítáme a pro $\alpha < p$-hodnota nezamítáme.

Námi pozorované hodnoty pro dané první signifikantní číslice k porovnání shody s Benfordovým zákonem jsou dány v tabulce níže.

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	$\sum$
youtube zhlédnutí	69	41	25	17	16	11	9	3	10	201
volby v MS kraji	191	107	65	57	55	49	34	41	27	626
ceny položek NSN	150	57	55	64	39	41	40	35	19	500
facebook komentáře	42	22	13	13	9	7	1	7	3	117
konstanty	115	64	30	28	28	24	11	17	18	335
vše dohromady	567	291	188	179	147	132	95	103	77	1779

TABULKA 7. Tabulka pozorovaných hodnot.

Youtube zhlédnutí

Zdroj pozorovaných hodnot: *www.youtube.cz*
Výsledek Pearsonova χ^2 testu dobré shody: S výslednou p-hodnotou $= 0.3703$ nulovou hypotézu nezamítáme.
Komentář: Pozorované hodnoty signifikantních číslic byly sepsány z počtu zhlédnutí na youtube. Hodnoty zhlédnutí jsme získali z hudebních videí různých hudebních skupin a zpěváků. Počet rostl od 12 tisíc zhlédnutí po 795 miliónů.

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$\rho_n(d_1) = [\%]$	34.32	20.40	12.44	8.46	7.96	5.47	4.48	1.49	4.98
$\mathbb{B}(D_1 = d_1) = [\%]$	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*

TABULKA 8. Tabulka relativních četností prvních signifikantních číslic v procentech, pro youtube zhlédnutí.

Pro zhlédnutí videi na youtube jsme určili i počet druhých signifikantních číslic, jelikož rozsah hodnot čísel, u kterých jsme pozorovali signifikantní číslice, je veliký. Neshoda by v tomto případě mohla působit jako protiváha v celkové shodě souboru čísel s Benfordovým zákonem.

Výsledek Pearsonova χ^2 testu dobré shody: S výslednou p-hodnotou $= 0.9149$ nulovou hypotézu nezamítáme.

d_2	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$\#(D_2 = d_2)$	30	19	25	21	18	20	18	18	19	13
$\rho_n(d_2) = [\%]$	14.93	9.45	12.44	1045	8.96	9.95	8.96	8.96	9.45	6.47
$\mathbb{B}(D_2 = d_2) = [\%]$	*11.97*	*11.39*	*10.88*	*10.43*	*10.03*	*9.67*	*9.34*	*9.04*	*8.76*	*8.50*

TABULKA 9. Tabulka relativních četností druhých signifikantních číslic v procentech, pro youtube zhlédnutí.

Volby v Moravskoslezském kraji

Zdroj pozorovaných hodnot: *www.volby.cz*
Výsledek Pearsonova χ^2 testu dobré shody: S výslednou p-hodnotou =0.522 nulovou hypotézu nezamítáme.
Komentář: Pozorované hodnoty signifikantních číslic byly sepsány z počtu volebních hlasů jednotlivým kandidátům ve volbách do Poslanecké sněmovny Parlamentu ČR. Volby se konaly ve dnech 25. a 26. října 2013 v Moravskoslezském kraji. Pro 626 kandidátů byla nejnižší hodnota počtu hlasů rovna 1 a ta nejvyšší 22090. V našem pozorování má toto samostatné téma největší p-hodnotu a zároveň největší počet pozorovaných hodnot.

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$\rho_n(d_1) = [\%]$	30.51	17.09	10.38	9.11	8.79	7.83	5.43	6.55	4.31
$\mathbb{B}(D_1 = d_1) = [\%]$	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*

TABULKA 10. Tabulka relativních četností prvních signifikantních číslic v procentech, pro volby v Moravskoslezském kraji.

Ceny položek NSN

Zdroj pozorovaných hodnot: *www.data.gov*
Výsledek Pearsonova χ^2 testu dobré shody: S výslednou p-hodnotou = $7.546 \cdot 10^{-4}$ nulovou hypotézu zamítáme.
Komentář: Pozorované hodnoty signifikantních číslic byly sepsány z cen položek označených skladovým číslem nato (**N**ational **S**tock **N**umber). Ceny byly vybrány náhodně z 9500 položek. Vybrané hodnoty nabyly nejníže ceny 0.41\$ a nejvýše 2245\$. I když nulovou hypotézu zamítáme, lze si všimnout sestupné frekvence výskytu vyšších signifikantních číslic.

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$\rho_n(d_1) = [\%]$	30	11.4	11	12.8	7.8	8.2	8	7	3.8
$\mathbb{B}(D_1 = d_1) = [\%]$	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*

TABULKA 11. Tabulka relativních četností prvních signifikantních číslic v procentech, pro ceny položek NSN.

Facebook komentáře

Zdroj pozorovaných hodnot: *www.facebook.com*
Výsledek Pearsonova χ^2 testu dobré shody: S výslednou p-hodnotou = 0.4279 nulovou hypotézu nezamítáme.
Komentář: Pozorované hodnoty signifikantních číslic byly sepsány z počtu komentářů jednotlivých příspěvků od různých osobností a facebook stránek. Nejnižší pozorovaný počet komentářů byl 62 a nejvyšší 101 781. I když relativní četnost pro sedmou signifikantní se dosti liší, nulová hypotéza není zamítnuta kvůli nízkému počtu pozorovaných hodnot.

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$\rho_n(d_1)$ = [%]	35.90	18.80	11.11	11.11	7.69	5.98	0.86	5.98	2.56
$\mathbb{B}(D_1 = d_1)$ = [%]	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*

TABULKA 12. Tabulka relativních četností prvních signifikantních číslic v procentech, pro Facebook komentáře.

Konstanty

Zdroj pozorovaných hodnot: *www.nist.gov/pml/*
Výsledek Pearsonova χ^2 testu dobré shody: S výslednou p-hodnotou = 0.2194 nulovou hypotézu nezamítáme.
Komentář: Pozorované hodnoty signifikantních číslic byly sepsány z hodnot fyzikálních a chemických konstant k roku 2014. Tento číselný soubor je častokrát zmiňovaný v souvislosti s Benfordovým zákonem. Fyzikální konstanty pravděpodobně vzbuzují dojem, že jsou dosti odlišné, a přitom se jejich rozdělení signifikantních číslic blíží Benfordovu zákonu, čili v tomto kontextu jsou zajímavé. Některé konstanty nabývají stejné hodnoty. Buďto je to jedna a tatáž konstanta pro různá vědecká odvětví nebo je to součin některé konstanty s konstantou 1 a tím pádem je to konstanta jiná. V tomto případě tedy mohla být frekvence signifikantních číslic ovlivněna právě těmito stejnými konstantami. Nicméně pořád vykazují vlastnost signifikantních číslic Benfordova zákona.

d_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$\rho_n(d_1)$ = [%]	34.33	19.10	8.96	8.36	8.36	7.16	3.28	5.08	5.37
$\mathbb{B}(D_1 = d_1)$ = [%]	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*

TABULKA 13. Tabulka relativních četností prvních signifikantních číslic v procentech, pro konstanty.

Vše dohromady

Výsledek Pearsonova χ^2 testu dobré shody: S výslednou p-hodnotou = 0.1085 nulovou hypotézu nezamítáme.
Komentář: Test dobré shody rozdělení signifikantních číslic všech předchozích pozorování.

d_i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$\rho_n(d_1)$ = [%]	31.87	16.36	10.57	10.06	8.26	7.42	5.34	5.79	4.33
$\mathbb{B}(D_1 = d_1)$ = [%]	*30.10*	*17.61*	*12.49*	*9.69*	*7.92*	*6.69*	*5.80*	*5.12*	*4.58*

TABULKA 14. Tabulka relativních četností prvních signifikantních číslic v procentech, pro všechny pozorované hodnoty.

Závěr

Ve své práci jsem ukázal pravděpodobnostní rozdělení signifikantních číslic, nadefinoval jsem benfordovské matematické struktury. Popsal jsem některé vlastnosti specifické pro Benfordův zákon, které dodávají Benfordovu zákonu na atraktivitě. Konvergencí posloupností náhodných veličin v distribuci k Benfordovu zákonu jsem se snažil o vysvětlení výskytu tohoto specifického pravděpodobnostního rozdělení signifikantních číslic. V poslední řadě jsem chtěl testováním ukázat, proč je vhodné se Benfordovým zákonem dále zabývat.

Ve všech kapitolách jsem se v podstatě snažil o co nejlepší popis a výklad vlastností Benfordova zákona z matematického hlediska. Účelem práce bylo udělat si obrázek o tomto fenoménu a dále volně aplikovat znalosti zde získané buď k rozšíření pole vlastností Benfordova zákona, nebo k analýze různých číselných souborů. Dále by z Benfordova zákona a jeho vlastností mohl vzniknout samostatný statistický nástroj. V poslední řadě by mohlo připadat v úvahu studovat číselné soubory jako výjimku Benfordova zákona.

Literatura

[BBH] BERGER, A., BUNIMOVICH, L., HILL, T.P., *One dimensional dynamical systems and Benford's Law* [online], *Trans. Amer. Math. Soc.* 357, s. 197-219, (2005), MR2098092, Dostupné z: *http://www.ams.org/journals/tran/2005-357-01/S0002-9947-04-03455-5/*

[Ben] BENFORD, F., *The Law of Anomalous Numbers* [online], *Proceedings of the American Philosophical Society,* Vol. 78, No. 4, s. 551-572, American Philosophical Society (Mar. 31, 1938), Dostupné z: *http://www.jstor.org/stable/984802*

[Ber] BERGER, A, *Multi-dimensional dynamical systems and Benford's Law* [online], *Discrete and continuous dynamical systems*, Vol. 13, No. 1, Website: http://aimSciences.org (2005), Dostupné z: *http://www.math.ualberta.ca/~aberger/papers/benford_multid_dcds.pdf*

[BH] BERGER, A., HILL, T.P., *A basic theory of Benford's Law* [online], *Probability Surveys*, Vol. 8, s. 1-126, (2011), DOI: 10.1214/11-PS175, Dostupné z: *http://www.i-journals.org/ps/viewissue.php?id=11#Articles*

[BL] BRIŠ, R., LITSCHMANNOVÁ, M., *STATISTIKA I. pro kombinované a distanční studium* [online], Vysoká škola báňská - Technická univerzita Ostrava, Fakulta elektrotechniky a informatiky Katedra aplikované matematiky, (2004), Dostupné z: *http://homel.vsb.cz/~bri10/*

[KN] KUIPERS, L., NIEDERREITER, H., *Uniform distribution of sequences* [online]. Mineola, N.Y.: Dover Publications, (2006), ISBN 04-864-5019-8, Dostupné z: *http://web.maths.unsw.edu.au/~josefdick/preprints/KuipersNied_book.pdf*

[Ne] NEWCOMB, S., *Note on the Frequency of Use of the Different Digits in Natural Numbers* [online], *American Journal of Mathematics,* Vol. 4, No. 1, s. 39-40, The Johns Hopkins University Press (1881), DOI: 10.2307/2369148, Dostupné z: *http://www.jstor.org/stable/2369148*

[NMMS] RIEČANOVÁ, Z., HORVÁTH, J., OLEJČEK, V., RIEČAN, B., VOLAUF, P., *Numerické metódy a matematická štatistika*, 1. vyd. Bratislava: Alfa, (1987). Edícia matematicko-fyzikálnej literatúry.

[Sm] SMITH, S.W., *Explaining Benford's Law* [online], Kapitola 34 z: *The Scientist and Engineer's Guide to Digital Signal Processing*, s. 701-722, California Technical Pub., (1997), Dostupné z: *http://www.dspguide.com/ch34.htm*

Příloha A

CD-ROM se softwarem.